(12) **EUROPEAN PATENT APPLICATION**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU**
**MC NL PT SE**
Designated Extension States:
**AL LT LV MK RO SI**

(71) Applicant: Geneart GmbH
93053 Regensburg (DE)

(72) Inventors:
• Wagner, Ralf
93049 Regensburg (DE)

• Graf, Markus
86316 Friedberg (DE)
• Deml, Ludwig
93128 Regenstauf (DE)
• Bieler, Kurt
93051 Regensburg (DE)

(74) Representative: Dehmel, Albrecht, Dr.
Dehmel & Bettenhausen
Patentanwälte,
Müllerstrasse 1
80469 München (DE)

(54) **Synthetic gagpol genes and their uses**

(57) The present invention relates to synthetic gag and gagpol genes optimized for high level expression via codon optimization and the uses thereof for the efficient generation of vector particles. The invention further relates to the generation of packaging cells and vaccines based on the synthetic gag and gagpol genes.

Fig. 1
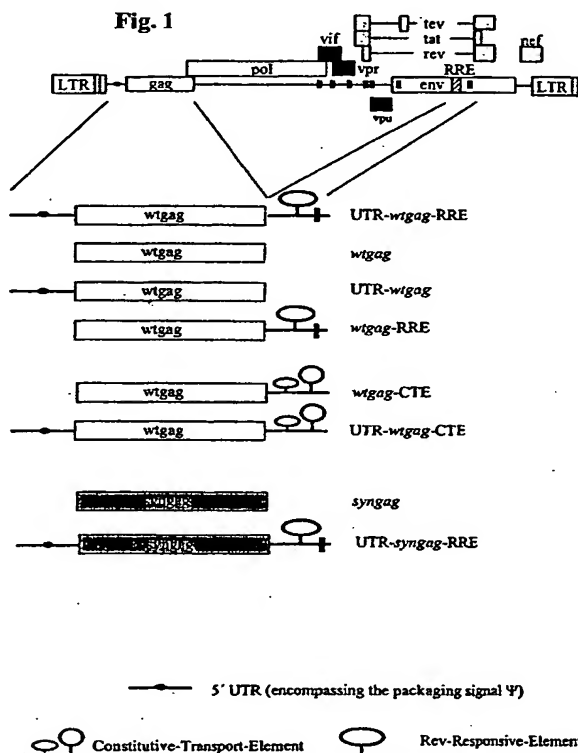
## Description

[0001] The present invention relates to synthetic *gag* and *gagpol* genes optimized for high level expression via codon optimization and the uses thereof for the efficient generation of vector particles. The invention further relates to the generation of packaging cells and vaccines based on the synthetic *gag* and *gagpol* genes.

[0002] Retroviral based vectors for DNA vaccination or gene therapy should require the ability to express structural genes and enzymatic functions in high yields in order to obtain maximal efficiency and yet, at the same time, with lowest possible risk and side effects for the recipient. Efficient production of structural or enzymatic proteins from complex retroviruses like HTLV-1 and -2, Lentiviruses and Spumaviruses is limited by their complex regulatory mechanism of transcription, nuclear RNA translocation and expression. For example the expression of SIV and HIV-1 late gene products is highly regulated both on transcriptional and post-transcriptional level therefore depending on the presence and function of early phase proteins like Tat and Rev. Whereas the anti-repressor function of Tat can be easily avoided by using heterologous transcriptional control via viral and/or cellular promotor/enhancer units, the expression of late genes, coding for structural or enzymatic functions (gag, env, pol) depends on the presence of the transactive Rev protein, known to promote the export of un- and partially spliced RNAs from the nucleus to the cytoplasm via the interaction with its cognate RNA recognition site Rev responsive-element (RRE), a complex of a 351 nucleotides (nt) long RNA stem-loop structure located within the env open-reading-frame. Although Rev/RRE action is well accepted as a necessary prerequisite for HIV-1 late gene expression, the critical contribution of different cis-acting elements within the un- and singly spliced transcripts to Rev dependency and timely regulated expression is still controversially discussed. Accordingly, Rev dependent and timely regulated export of late singly or unspliced lentiviral mRNAs has been explained in most reports either by inefficient splice site formation or attributed to inhibitory sequences located within the coding region (referred to as INS elements).

[0003] Whereas the exact nature and function of so called INS elements is presently not completely clear, Rev dependent nuclear export and expression of the un- and singly spliced lentiviral mRNAs, at least when expressed from heterologous promotors, seems to engage and require the major splice donor site located in the untranslated region 5' of the Gag coding sequence (5'-UTR). This 5'-UTR also contains part of the RNA packaging signal (ψ), which - in contrast to the ψ-sites of traditional C and D-type retroviruses - also extends into the coding area of lentiviral genomes including the Gag and Pol reading frame.

[0004] Although lentiviral vectors derived from complex retroviruses like HTLV-1 and -2, Lentiviruses and Spumaviruses offer great promise in the field of DNA vaccination and/or gene therapy, concerns regarding safety in humans still exist, due to the striking pathogenic potential of HTLV's and lentiviruses like HIV-1 or HIV-2. Moreover, due to the complex viral regulatory mechanisms underlying Gag and GagPol expression efficient expression is either limited by or depends heavily on the presence of cis-acting elements (UTR, RRE) and transactive proteins (Rev, Tat). Tat-independent transcription can be easily achieved by employing a constitutive mammalian promotor like the Cytomegalovirus immediate early promotor/enhancer unit. Rev independent expression of late HIV-1 genes such as those coding for structural or enzymatic proteins would be highly appreciable in order to improve efficiency and safety of lentiviral based gene therapy vectors and DNA vaccines as well as for antigen production in a mammalian expression system. Generally, two systems can be applied: (a) The Mason-Pfitzer monkey virus (MPMV) constitutive transport element (CTE), a cis-acting RNA element located within the 3' untranslated region (UTR) of the viral genome, can substitute the HIV-1 Rev/RRE regulatory system. In addition, several other cis-acting RNA elements of a variety of viruses also promoting nuclear export of intron containing RNAs were discovered (e.g. Rous sarcoma virus, Simian retrovirus type D, Avian leukemia virus, HBV). Accordingly, the addition of the MPMV-CTE element or functional analogous cis-acting RNA elements overcomes the effect of the so called INS elements, which have been proposed to account for the nuclear sequestration of late lentiviral RNAs in the absence of Rev. (b) Low or no gene expression of late HIV-1 gene products in absence of Rev has been partially overcome by clustered single point mutations within the wobble positions of selected codons within coding DNA sequences (Schwartz *et al.* 1992a J. Virol. 66:7176-7182; Schneider *et al.* 1997, J. Virol 71:4892-4903; Haas *et al.* 1996, Curr. Biol.: 6:315-324). This approach aimed to destroy the so called INS elements in order to render the expression of the resulting Gag gene independent of Rev. However, the presence, nature and function of INS elements is still discussed controversially, indicating that a complete knockout of poorly characterized INS elements on the basis of single point mutations may be difficult to achieve. This view is confirmed by the observation of Qiu and colleagues (Qiu *et al.* 1999 J Virol. Nov;73(11):9145-52.), showing that the addition of a CTE element to the (only partially) altered Gag gene leads to a further increase of Gag expression. This phenomenon does not hold true for a completely Rev-independent Gag expression construct.

[0005] A preferred application of HTLV-1 and -2 as well as Lentivirus derived Gag and GagPol based expression vectors is DNA vaccination and the production of antigens, Preferably, virus-like particles, in higher eucaryotic expression systems like in mammalian and insect cells. Cell mediated immune responses to Gag and Pol products of HIV-1 were shown to protect from disease or at least to contribute to an efficient control of virus replication. Accordingly, in persons with chronic infection, HIV-1-specific proliferative responses to p24 as well as the number of Gag specific CTL

precursors were inversely related to the determined viral loads. Moreover, in CTL responses directed towards highly conserved epitopes within Gag or Pol seem to critically contribute to the control of virus replication in long-term non-progressing HIV infected individuals.

[0006] Different formats have been devised in the past towards presenting Gag and Pol containing immunogens to the immune system, recombinant virus-like particles (VLPs) and Gag/Pol containing DNA constructs being amongst them. VLPs were usually produced in a baculovirus driven insect cell expression system, which allows to bypass the complex regulation of Gag and Pol expression seen in the natural host or mammalian cells. Such VLPs turned out to be highly immunogenic in rodents as well as in nonhuman primates. However, to compile with the regulations of Good Manufacturing Practice (GMP) and in order to achieve proper posttranslational modification of Gag/Pol polypeptides (including coexpressed proteins that might be packaged in or presented by VLPs), expression of GagPol and derivatives thereof in mammalian cell lines would be highly appreciated.

[0007] Vaccination by direct injection of DNA is a novel and promising method for delivering specific antigens for immunization. Plasmid DNA immunization has potential advantages compared to traditional protein vaccination due to the strong T helper 1 (Th1) and CTL responses induced, the prolonged antigen expression and the long lived effector activity, and thus can be used for vaccination. In animal models, DNA vaccination has been shown to induce protective immunity against a variety of viral and parasitic pathogens. In most cases, strong, yet highly variable, antibody and cytotoxic T cell responses were associated with control of infection. Plasmid DNAs expressing genes derived from HIV-1 have been recently shown to induce humoral and cellular immune responses in rodents, in nonhuman primates and in phase I studies in humans. Although these constructs were able to induce an immune response, both circulating antibody titers and HIV-1 specific CTL levels were transient and low.

[0008] However, the development of DNA constructs and infectious, although not necessarily replication-competent, vectors expressing Gag or GagPol or derivatives thereof for antigen production in eucaryotic cells and for vaccination purposes faces several limitations both, regarding safety and efficiency. Gag and GagPol expression using *wildtype* genes in the Rev-dependent situation is limited (a) by the complex viral regulatory mechanisms which involves several cis-acting elements (RRE, UTR) and trans-acting proteins (Rev, Tat). In absence of UTR, RRE, or Rev no or only minute amounts of Gag or GagPol protein will be produced. Gag or Gag/Pol expression therefore needs simultaneous expression of the Rev protein either from a bicistronic construct or from a separate plasmid. Both strategies limit the efficiency of generating cell lines in vitro or reduce the efficacy of DNA vaccine constructs *in vivo* either due to an increase plasmid size or the necessity to transfect/transduce one single cell with both plasmids at the same time; and (b) by the presence of the Rev protein itself, acting as an RNA shuttle between nucleus and therefore harboring an intrinsic risk. Moreover, when applied as a therapeutically vaccination in chronically HIV infected individuals, Rev could contribute towards reactivating latent viruses, thereby enhancing the infection, virus replication and disease.

[0009] Moreover, Gag and GagPol expression - irrespective of accomplished by the Rev/RRE system or mediated independent from Rev/RRE by addition of a constitutive transport element (CTE) or analogous cis acting sites - is limited by the presence of UTR, RRE and several stretches in the GagPol coding area, which are known to comprise the HIV-1 packaging signal ψ or other RNA elements responsible for packaging the genomic RNA into viral particles. Vectors comprising such RNA elements could be packaged into and distributed by endogenous retroviral particles and contribute towards the mobilization of endogenous retroviruses. Additionally, Gag or GagPol derived VLPs generated either in cell culture or *in vivo* following DNA vaccination would self package their own RNA, which precludes any safe application of such DNA vaccine constructs for vaccination purposes in humans.

[0010] Moreover, any kind of Gag and GagPol expression - irrespective of being achieved (i) by Rev/RRE system, (ii) by addition of CTE or analogous sites or (iii) after introduction of single point mutations in the Gag or GagPol gene - is limited by the wildtype sequence itself in the case of a Gag or GagPol DNA vaccine. Recombination events could occur between the vaccine construct itself (UTR, RRE, wildtype ORF) and endogenous retroviral sequences or, when administered as a therapeutic vaccine, between the nucleic acids of the vaccine construct and the genetic information of the virus circulating within the patient. Both could lead to the reconstitution of potentially infectious or chimeric particles inheriting unknown risks.

[0011] Another preferred application of lentiviral Gag and GagPol based vectors is gene transfer into different types of dividing, resting or postmitotic cells. In contrast to standard retroviral gene transfer as mediated e.g. by Moloney Murine Leukemia Virus (MoMuLV) based vectors, lentiviral gene transfer can be used for gene delivery into a variety of quiescent or non-dividing cells, such as neuronal, muscle, and liver as well as hemapoietic stem cells and might therefore considerably contribute to the prevention and treatment of genetic disorders, tumor diseases and infections. The transduction of non-dividing, resting or postmitotic cells as neuronal, hepatical tissue or hematopoetic progenitor cells by for the treatment of variety of genetic disorders or acquired diseases. The lentiviral vector particles are currently prepared by triple-transfection of a *GagPol* expression plasmid together with a transfer construct (containing the packaging signal and transfer gene; flanked by LTRs), and expression plasmid encoding an envelope protein derived from an amphotropic or xenotropic retrovirus like Moloney Murine Leukemia Virus or Vesicular stomatitis virus into mammalian cell lines. Lentiviral vector particles from the supernatant of such transfected cells were able to stable transduce

3

a large variety of different cells even after *in vivo* gene transfer. To increase the safety of these vectors, most of the accessory genes of HIV-1 were deleted from the packaging construct and the vector, thereby minimizing the risk for the emergence of potentially pathogenic replication competent recombinants.

[0012]   However, the development of lentiviral Gag and GagPol based vectors for gene transfer into quiescent or non-dividing cells faces several limitations regarding safety and efficiency.

[0013]   The currently used HIV-1 or SIV derived *GagPol* expression plasmids contain parts of the 5'-untranslated region comprising the RNA packaging signal ψ. Although small deletions were introduced into the assumed packaging signal, it is unlikely that these deletions completely prevent packaging, since similar deletions in the context of an intact 5'leader sequence reduced packaging by less than a factor of 25 (Luban and Goff, 1994 J Virol. Jun;68(6):3784-93).

[0014]   Moreover, if the *GagPol* expression plasmid can indeed be packaged by homologous or non-homologous recombination events between the vector RNA and the *GagPol* RNA during or after reverse transcription can not be excluded. This could lead to the - for safety reasons - undesired transfer of the *GagPol* gene (or parts of it) into target cells.

[0015]   Moreover, there are two regions of homology between the *GagPol* expression plasmids and the lentiviral transgene vector constructs: (1) Rev-dependent as well as CTE mediated, Rev independent GagPol expression (*packaging* proteins) requires either the complete UTR or at least part of it for efficient expression. Also the *transgene* vector constructs depend for efficient packaging on the RNA packaging signal that overlaps the 5'-UTR. (2) Since in HIV-1, SIV and other lentiviruses the packaging signal extends from the UTR into the *gag* gene (reviewed in Swanstrom-R and Wills-JW, 1997, Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 263-334.), the 5'-portion of the *gag* gene is part of many lentivirus derived transgene vector constructs. (3) Finally both, the transgene vector construct and Rev-dependent *GagPol* expression plasmids contain the Rev-responsive element, since (i) RRE was suggested to include RNA packaging functions and (ii) the transport of the transgene vector RNA and the *GagPol* RNA from the nucleus to the cytoplasm is Rev-dependent. These regions of extensive homology might facilitate homologous recombination events during either vector production or reverse transcription.

[0016]   The problem underlying the present invention is the development of retroviral based expression vectors allowing high level protein production in absence of transactive proteins and with lowest risk of self packaging of the vector or recombination events thereby avoiding the risk for reconstitution of potentially infectious or by other means hazardous particles. A further problem underlying the present invention is the provision of vaccines against diseases caused by Lentiviruses.

[0017]   The problem of the invention is solved by the subject matter defined in the claims.

[0018]   The present invention is further illustrated by the figures.

FIG. 1:

[0019]   Schematic representation of wildtype and synthetic gag encoding expression plasmids. Open boxes indicate wildtype gag (wtgag) encoding genes including 3' and 5' located cis-acting sequences, whereas syngag without any 3' or 5' untranslated regions (UTR) is boxed black. Wtgag reading frames were fused to the cis-acting sequences 5' located UTR, the Rev responsive-element (RRE), the constitutive transport element (CTE), or combination thereof. To investigate the influence of the major splice donor (SD) on CTE mediated Rev-independent Gag expression, the UTR were mutated in order to destroy the splice-donor consensus sequence within this upstream located sequence, resulting in mutant ΔSD-wtgag-CTE.

FIG. 2:

[0020]   Representation of syngag and wildtype gag open reading frames. (A) The Gag coding sequence was adapted to a codon usage occurring in highly expressed mammalian genes and aligned to the corresponding wildtype sequence shown above. Sequence identity between the synthetic and wildtype gene is indicated by lines. (B) The table indicates the different sequence composition of the wildtype and synthetic gag reading frame. The fully synthetic gag gene with an optimized codon usage shows a markedly increase in G/C content whereas the overall A/U content is reduced. Moreover codon usage adaptation increased the amount of immune-stimmulatory CpG islets (Pur-Pur-C-G-Pyr-Pyr) within the coding region from only 1 to 5.

FIG. 3:

[0021]   H1299 cells were transiently transfected with indicated plasmids. For standardization Gag expression was compared to wildtype like Rev/RRE dependent expression by a Rev expression vector, whereas Rev-independent expression plasmids were co-transfected with vector (indicated as + or - respectively). (A) Rev/RRE mediated wildtype expression. (B) CTE-mediated wildtype expression. (C) Synthetic expression using a codon optimized reading frame.

Cells were harvested 48 hours post-transfection, lysed and 50 μg of total protein subjected to Western Blot analysis (A-C, lower panel). Yields of Pr55$^{gag}$ were measured by testing different dilutions of cell-lysate in a capture-ELISA using purified Pr55$^{gag}$ for standardization (A-C, upper panel) and were normalized by the Pr55$^{gag}$ production obtained by autologous Rev/RRE dependent expression. Gag-expression levels obtained by transfection of different cell lines were performed to rule out cell type specific effects (D). Bars represent relative Pr55$^{gag}$ expression levels (taken UTR-wtgag-RRE+Rev as 100%) and are given as the mean of triplicate determinations.

FIG. 4:

[0022]    Immunological analysis of an HIV-1 DNA Gag vaccine based on synthetic and wildtype sequences. 80 μg of wildtype (wt-gag) and synthetic (syngag) encoding DNA expression plasmids were injected intramuscularily (im) into BALB/c mice and boosted twice, after 3 weeks and 6 weeks and sacrificed after 7 weeks. (A) Strength of humoral responses induced in immunized BALB/c. Each bar represents the group mean (n=4) for anti-gag IgG1, anti-gag IgG2 and total IgG as determined by end-point dilution ELISA assay. End-point titers of the immune sera were defined as the reciprocal of the highest plasma dilution that resulted in an adsorbance value (OD 492) three times greater than that of a preimmune serum with a cut-off value of 0,05. (B) Cytotoxic T cell activity in splenocytes from mice immunized intramuscularly or subcutaneously by particle gun (p.g.) with gag expression vectors. Lymphoid cells obtained from mice five days after the boost injection were co-cultured with gag peptide-pulsed syngenic P815 mastocytoma cells (irradiated with 20,000 rad). Control assays included splenocytes of non immunized and wt-gag immunized mice stimulated *in vitro* with peptide-pulsed P815 cells. Cytotoxic effector populations were harvested after 5 days of *in vitro* culture. The cytotoxic response was read against gag-peptide pulsed A20 cells and untreated A20 negative target cells in a standard $^{51}$Cr release assay. Data shown were mean values of triplicate cultures.

FIG. 5:

[0023]    Maps of the (B) SIV-derived packaging constructs based on optimized GagPol reading frames of SIV and HIV and (A) (A) and HIV-1 SIV-derived (B) transgene vectors or packaging constructs. Regions deleted from the SIV genome are marked by shaded boxes with the first and last deleted nucleotide given after the "Δ" sign (numbering according to Genbank entry M33262). Point mutations that inactivate reading frames are marked by asterisks, inactivated reading frames are marked in black. Synthetic GagPol reading frames are hatched. BGHpAd: bovine growth hormone polyadenylation signal; CMV: immediate early promoter/enhancer region from human cytomegalovirus; SFFV-Pr: Spleen focus forming virus U3 region; GFP: gene for the green fluorescence protein;

FIG. 6:

[0024]    Immunoblot analysis of cell culture supernatants harvested 3 days after transfection of H1299 cells with the indicated lentiviral (HIV-1: Hgp$^{syn}$ and SIV$_{mac239}$: Sgp$^{syn}$) GagPol expression constructs and proviral DNAs. Released GagPol particles were enriched by continuous sucrose density gradient sedimentation. Antigen peak fractions were separated on a denaturing 12,5% SDS-PAGE and analyzed by immunobloting using HIV-1 p24 (A) or SIV p27 specific (B) monoclonal antibodies. Protease activity and functionality in the polyprotein processing was shown by absence (-) or presence (+) of Saquinavir. Molecular weights of the detected precursor proteins (Pr) and mature cleavage products are indicated. (C) Additionally, antigen peak fractions were also tested for its content of HIV-1 p24 capsid antigen using a commercially available capture ELISA format (DuPont, Boston, USA)

FIG. 7:

[0025]    Transduction of growth-arrested cells. 293T cells were transfected with ViGΔBH, pHIT-G and the indicated lentiviral *GagPol* expression plasmids. The MLV vector was generated by transfecting plasmids pLEGFP-N1, pHIT60, and pHIT-G. Vector titers in the supernatant of transfected cells were determined on 293 cells in the presence of the indicated concentrations of aphidicolin. The titration was done in triplicates or quadruplicates. The means and the standard deviations are shown. GFU: green fluorescence forming units.

FIG. 8:

[0026]    Detection of replication competent recombinants. CEMx174-SIV-SEAP cells were infected with the supernatant of 293T cells co-transfected with SIV-GFPΔDP (A) or SIV-GFPΔBP (B) and the indicated *gag-pol* expression plasmids. RLU: relative light units.

5

FIG. 9:

[0027] Schematical representation of a codon optimized GagPol derivative which can be used for DNA vaccination. For safety and regulatory reasons the packaging signal Ψ was removed, the integrase deleted, and the reverse transcriptase (RT) gene disrupted by insertion of a scrambled nef gene at the 3'end of the DNA sequence coding for the RT active site. The nef gene was dislocated by fusing its 5'half to its 3'half. Myristilation is of the GagPolNef particles is inhibited by a Glycin to Alanin mutation at the myristilation site of the gag gene.

[0028] The term "packaging proteins" as used herein refers to GagPol proteins of retroviruses, spumaviruses, Preferably, HTLV-1 and 2, more Preferably, all lentiviruses, most Preferably, HIV-1, HIV-2, SIV, FIV, and in particular genes encoded by SEQ ID NO:2 (GagPol HIV-1$_{IIIB}$) and SEQ ID NO:1 (GagPol SIV$_{mac239}$) which are capable of packaging an RNA containing the packaging site Ψ such as an retroviral, Preferably, lentiviral, most Preferably, HIV-1, HIV-2, SIN, FIV genome and/or a so called "transfer construct" which is specified in detail below.

[0029] The term "packaging construct" as used herein refers to an expression plasmid or vector system used to produce packaging proteins in eucaryotic cells.

[0030] The term "transfer construct" as used herein refers to an expression plasmid or vector system used to produce an RNA containing the packaging site Ψ, the transfer gene and is flanked by LTRs. This RNA is capable to be incorporated within an transducing particle which itself is produced by an eucaryotic cell expressing packaging proteins. The LTRs and the packaging site Ψ is of retroviral, Preferably, spumaviral or lentiviral and most Preferably, of HIV-1, HIV-2, SIV or FIV origin.

[0031] The term "vector particle" as used herein refers to a transduction competent viral particle comprising packaging proteins, envelope proteins as exposed on the surface of the viral particle as well as an incorporated transfer construct.

[0032] The present invention relates to nucleic acid sequences encoding the gag and pol polypeptides, whereby the amino acid Ala is encoded by the nucleic acid triplett gcc or gct, Arg is encoded by agg or aga, Asn is encoded by aac or aat, Asp is encoded by gac or gat, Cys is encoded by tgc or tgt, Gln is encoded by cag or cat, Glu is encoded by gag or gaa, Gly is encoded by ggc or gga, His is encoded by cac or cat, Ile is encoded by atc or att, Leu is encoded by ctg or ctc, Lys is encoded by aag or aat, Met is encoded by atg, Phe is encoded by ttc or ttt, Pro is encoded by ccc or cct, Ser is encoded by agc or tcc, Thr is encoded by acc or aca, Trp is encoded by tgg, Tyr is encoded by tac or tat, Val is encoded by gtg ot gtc and the stop codon is tga or taa.

[0033] Preferably, the present invention relates to nucleic acid sequences encoding the gag and pol polypeptides, whereby the amino acid Ala is encoded by the nucleic acid triplett gcc or gct, Arg is encoded by agg or aga, Asn is encoded by aac or aat, Asp is encoded by gac or gat, Cys is encoded by tgc or tgt, Gln is encoded by cag or cat, Glu is encoded by gag or gaa, Gly is encoded by ggc or gga, His is encoded by cac or cat, Ile is encoded by atc or att, Leu is encoded by ctg or ctc, Lys is encoded by aag or aat, Met is encoded by atg, Phe is encoded by ttc or ttt, Pro is encoded by ccc or cct, Ser is encoded by agc or tcc, Thr is encoded by acc or aca, Trp is encoded by tgg, Tyr is encoded by tac or tat, Val is encoded by gtg ot gtc and the stop codon is tga or taa, whereby the nucleic acid sequence in the region in which the reading frames encoding the gag and pol polypeptides overlap corresponds to the wildtyp nucleic acid sequence encoding the gag and pol polypeptides.

[0034] Most Preferably, the present invention relates to nucleic acid sequences as depicted in SEQ ID NO:1 or 2.

[0035] We were able to demonstrate that consequent and strict codon optimization of the Gag and GagPol reading frames derived from the Simian Immunodeficiency virus SIV$_{mac239}$ or the Human Immunodeficiency virus HIV-1$_{IIIB}$ allowed high level expression of Gag, GagPol and derivatives thereof in complete absence of Rev/RRE, CTE and UTR which can be used directly for DNA vaccination.

[0036] Moreover, we were able to demonstrate that consequent and strict codon optimization of SIV$_{mac239}$ or HIV-1$_{IIIB}$ GagPol reading frames with the exception where Gag and Pol reading frames are overlapping, allowed high level expression of GagPol and derivatives thereof in complete absence of Rev/RRE, CTE and UTR and can be used as packaging constructs for gene delivery.

[0037] Therefore, *safe* and *efficient* delivery and expression of Gag or GagPol genes according to the present invention allows (i) expression of high yields of native proteins in higher eucaryotes like in mammalian or insect cells and (ii) supporting the generation of transduction competent particles for vaccination and gene therapy purposes which is currently limited by the complex regulatory mechanisms of late Gag and GagPol expression.

[0038] The present invention is based on the observation, that the codon usage in the HIV genome (A/T content 55.88 %) significantly deviates from that of highly expressed mammalian genes (usually not exceeding 50%) indicating a regulatory function of the codon usage for timely regulated lentiviral gene expression.

[0039] Additionally, we were able to show that CTE-mediated, similar to the Rev/RRE mediated, expression of wildtype HIV-1 late gene products (such as gag), essentially depends on the presence of the HIV-1 5'-untranslated region (UTR) or at least parts of it, which also contains an RNA stretch responsible for packaging of the genomic RNA into the viral particle, and therefore touches sensible safety issues.

[0040] In order to eliminate inhibitory or otherwise regulatory sequences within lentiviral Gag and GagPol genes, reading frames were optimized according to codon usage found in highly expressed mammalian genes. Preferably,, codons that are frequently found in highly expressed mammalian genes (Ausubel *et al.* 1994 *Current Protocols in Molecular Biology* 2, A1.8-A1.9). most Preferably, in highly expressed human genes were used to create a fully synthetic reading frame not occurring in nature, which, however encodes the very same product as the original wildtype gene construct, For that purpose, a matrix was generated considering almost exclusively those codons that are used most frequently and, less Preferably,, those that are used second most frequently in highly expressed mammalian genes as depicted in table 1.

Table 1:

| Codon most frequently (codon 1) and second most frequently (codon 2) found in highly expressed mammalian genes. | | |
|---|---|---|
| amino acid | codon 1 | codon 2 |
| Ala | GCC | GCT |
| Arg | AGG | AGA |
| Asn | AAC | AAT |
| Asp | GAC | GAT |
| Cys | TGC | TGT |
| End | TGA | TAA |
| Gln | CAG | CAT |
| Glu | GAG | GAA |
| Gly | GGC | GGA |
| His | CAC | CAT |
| Ile | ATC | ATT |
| Leu | CTG | CTC |
| Lys | AAG | AAT |
| Met | ATG | ATG |
| Phe | TTC | TTT |
| Pro | CCC | CCT |
| Ser | AGC | TCC |
| Thr | ACC | ACA |
| Trp | TGG | TGG |
| Tyr | TAC | TAT |
| Val | GTG | GTC |
| (Ausubel *et al.* 1994 *Current Protocols in Molecular Biology* 2, A1.8-A1.9). | | |

[0041] Based on this matrix, the Wisconsin *genetics computer group* (gcg) software package was employed to create the synthetic reading frame by backtranslating the amino acid sequence of lentiviral Gag and GagPol polypeptides into the codon optimized synthetic reading frame.

[0042] Few deviations from strict adherence to the usage of most frequently found codons may be made (i) to accommodate the introduction of unique restriction sites at approximately 250 to 300 nt intervals (see e.g. Fig. 2) (ii) to break G or C stretches extending more than 7 base pairs in order to allow consecutive PCR amplification and sequencing of the synthetic gene product. This may also be performed for the Gag and Pol based expression plasmids used for DNA vaccination as well the packaging constructs used for gene transfer with exception of the latter where *gag* and *pol* reading frames are overlapping. In particular it is preferred that the overlapping region of the *gag* and *pol* reading frames of the packaging constructs used for gene transfer may not be altered at position 1298 - 1558 of SEQ ID NO: 2 and position 1175 - 1532 of SEQ ID NO:1, in order to maintain both reading frames.

**[0043]** In particular, the use of synthetic Gag and Pol genes comprising the GagPol genes of all Retroviruses, Human T-cell Leukemia Virus 1 and 2, Spuma- and Preferably, Lentiviruses, in particular derived from HIV and SIV as encoded by SEQ ID NO:2 (GagPol HIV-1$_{IIIB}$) and SEQ ID NO:1 (GagPol SIV$_{mac239}$) allows therefore the safe and efficient production of packaging proteins for delivering genetic information to cells, whereas codons of Gag and GagPol reading frames were adapted to highly expressed mammalian genes as specified above, except regions where gag and pol reading frames are overlapping.

**[0044]** The present invention allows therefore the safe and efficient production of packaging proteins by increasing the yields of GagPol synthesis as compared to the expression rates achieved from the wildtype gene or wildtype like genes characterized by clustered point mutations - if driven by the same cellular or viral promotor/enhancer unit - as a result of consequent adaptation of the codon usage to that of frequently expressed mammalian genes.

**[0045]** The present invention allows the safe and efficient production of packaging proteins in complete absence of known cis-acting elements (UTR, RRE) or transactive proteins such as Rev and Tat.

**[0046]** The present invention further relates to retroviral gag or gagpol based particles generated by nucleic acid molecules comprising the nucleic acid sequence according to the invention having at least a 25 fold reduced incorporation rate of said nucleic acid molecules encoding the gag and pol polypeptides compared to retroviral particles generated by wildtyp nucleic acid sequences encoding the gag and pol polypeptides.

**[0047]** The present invention further relates to retroviral gag or gagpol based particles produced by nucleic acid molecules comprising the nucleic acid sequence according to the invention having at least a 100 fold reduced recombination rate between said nucleic acid molecules encoding the gag and pol polypeptides and a wildtype genome based transgene construct derived from the same or another retrovirus.

**[0048]** The present invention allows the efficient production of safe vector particles by reducing the opportunities for reconstituting infection competent hybrid retroviral particles by recombination events between nucleic acids encoding the packaging proteins and sequences encoding endogenous retroviruses.

**[0049]** The present invention allows the efficient production of safe vector particles by reducing the risk for incorporating the RNA species encoding the packaging proteins into vector particles by more than 25 fold as compared to expression constructs containing the 5'-UTR or encoding the wildtype GagPol gene or by diminishing the risk of packaging the very same RNA into endogenous retroviral particles.

**[0050]** The present invention allows the efficient production of safe vector particles by reducing the risk of recombination between the packaging construct and the gene transfer construct by more than 100 fold, by combining a synthetic gene encoding the GagPol packaging function from one retro- spuma or lentivirus with the wildtype genome based transgene construct derived from (i) the very same or (ii) another retro-, spuma- or lenti- virus.

**[0051]** More specifically, the present invention allows the efficient production of safe vector particles by combining the synthetic gene encoding the HIV-1 derived packaging proteins with the wildtype HIV-1 genome based transgene construct or by combining the synthetic gene encoding the SIV$_{mac239}$ derived packaging proteins with the wildtype SIV$_{mac239}$ genome based transgene construct or combinations thereof.

**[0052]** Preferably,, by combining the synthetic gene encoding the HIV-1 derived packaging proteins with the wildtype SIV$_{mac239}$ genome based transgene construct or by combining the synthetic gene encoding the SIV$_{mac239}$ derived packaging proteins with the wildtype HIV-1 genome based transgene construct, thereby reducing homologies between transfer vector and packaging vector (as e.g. depicted in table 1 of example 4) resulting in safe and efficient packaging constructs being useful for the production of safe lentiviral vectors and gene delivery systems.

**[0053]** Synthetic GagPol genes may or may not encode the active integrase. In the latter case, the integrase gene may be deleted in total. Alternatively, the enzyme activity may by knocked out by one deletion or several short continuous in frame deletions of various length, Preferably, not exceeding 5 amino acids and, more Preferably,, comprising a single codon. Also, the integrase may be rendered inactive by point mutations at various sites, Preferably, by a single point mutation (e.g. position 3930 - 3932 in SEQ ID No:2 and position 3957 - 3959 in SEQ ID No:1 are changed to glutamin encoding nucleotides).

**[0054]** The present invention furthermore relates to retroviral packaging cells for the generation of the retroviral particles according to the invention transformed with nucleic acid molecules comprising the nucleic acid sequence according to the invention.

**[0055]** GagPol packaging proteins according to the present invention may be stably, inducibly or transiently expressed in any primary or lab adapted cell or cell line of human, mammalian or non-mammalian origin. The synthetic GagPol genes may be episomal or chromosomal within these cells and may contain additional recombinant genes, preferable any specific fusogenic gene, more preferable any viral envelope gene, more preferable a retroviral envelope gene. These cells may also contain either an episomal or chromosomal transfer vector, but may not contain any additional viral wildtype sequences.

**[0056]** GagPol expression may be driven by any tissue- or cell type-specific promotor/enhancer such as muscle kreatin kinase or MHC class I promotor/enhancer or by any viral e.g. CMV immediate early, Rous Sarcoma Virus promotor/enhancer, Adenovirus major/late promotor/enhancer or nonviral e.g. synthetic promotor/enhancer.

[0057] GagPol expression may be either constitutive or inducible e.g. by addition of a specific inducer like ecdysone in the case of a hormone inducible promotor enhancer or by removal of a repressor protein (Tet on/off) or repressor gene (Cre/Lox).

[0058] GagPol based gene delivery into cells may be mediated by any means of transfecting or introducing a plasmid or plasmid derivative such as "midges" (closed linear DNA) encoding the synthetic GagPol gene or the above variants thereof into cells.

[0059] GagPol based gene delivery into cells may be also mediated by infectious recombinant (i) viral vectors such as recombinant Retroviruses, Vaccinia- or Poxviruses, Adenoviruses, Alphaviruses, Herpesviruses, Baculoviruses or any other recombinant virus, (ii) subviral components bridging transfection and infection procedures like e.g. Virosomes comprising nucleic acid/protein aggregates containing e.g. Influenca hemagglutinin, (iii) bacterial vectors such as recombinant Listeriae or Salmonellae or any other type of infectious, replicating or non-replicating vector.

[0060] Any means of introducing GagPol genes and derivatives thereof into cells may lead to a transient expression of GagPol. The introduction of the above indicated GagPol genes into cells may also allow the establishment of stable cell lines supporting either constitutive or inducible GagPol expression.

[0061] Retroviral, Preferably, Spumaviral and Lentiviral and, more specifically, HIV or SIV derived vector particles carrying the transgene RNA and being capable of transducing nondividing, resting or postmitotic cells of various species in addition to dividing or proliferating cells may be generated by co-expressing a synthetic GagPol constructs according to the invention together with any appropriate amphotropic envelope like the vesticular stomatitis virus G protein, xenotropic envelope like the env protein of avian or murine sarcoma viruses e.g. the Rous sarcoma virus (RSV) Env protein, or cell type specific receptor protein like the HIV-1 Env protein, in the presence of any appropriate gene transfer construct encoding the desired packaging competent transgene RNA.

[0062] Transduction competent vector particles may be generated by co-transfecting or co-infecting or transfecting/infecting primary cells or cell lines of various origin such as mammalian Verocells, HeLa, Cos, CHO or H1299 cells or insect SF9, High-5 or DS-2 cells with vector constructs encoding the synGagPol derived packaging proteins, and any appropriate envelope or receptor structure as well as the desired transgene.

[0063] Alternatively, transduction competent vector particles may be generated by stable cell lines expressing GagPol from the synthetic GagPol gene according to the invention, either inducible or constitutively, whereby the stable cell lines may be co-transfected or co-infected or transfected/infected with vector constructs encoding any envelope or receptor structure as well as the desired transgene.

[0064] Preferably,, transduction competent vector particles may be generated by stable cell lines expressing GagPol from the synthetic GagPol according to the invention gene together with any appropriate envelope or receptor structure, either inducible or constitutively, whereby the stable cell lines may be transfected or infected with vector constructs encoding the desired transgene.

[0065] Ideally, transduction competent vector particles may be obtained from a stable cell line mediating the expression of the synthetic GagPol gene and any appropriate receptor or envelope function together with the generation of the packaging competent transgene RNA, either in an inducible manner or constitutively or a combination thereof.

[0066] Most Preferably,, transduction competent vector particles are generated by transfection of cells with a plasmid or derivative thereof or by infecting cells with any infectious, although not necessarily replication competent vector encoding the synGagPol derived packaging proteins, and any appropriate envelope or receptor structure as well as the desired transgene.

[0067] Both transduction competent vector particles, as well as infectious, although not necessarily replication competent vectors encoding the synGagPol derived packaging proteins, and any appropriate envelope or receptor structure as well as the desired transgene may be used for application ex vivo and in vivo.

[0068] Furthermore, the present invention relates to nucleic acid molecules comprising the nucleic acid sequence according to the invention for use as an active pharmaceutical substance.

[0069] The present invention relates further to the use of nucleic acid molecules comprising the nucleic acid sequence according to the invention for the preparation of a pharmaceutical composition for the treatment and prophylactic of diseases caused by retroviruses.

[0070] Moreover, the present invention can be used for vaccination purposes, in particular, if used as nucleic acid based vaccination constructs, which have been shown to elicit a strong humoral and Th-1 like cellular immune response in mice. On the basis of the GagPol genes of the present invention it is now possible to construct effective retroviral, HTLV-1 and -2 derived as well as lentiviral and in particular HIV and SIV based Gag and GagPol vaccines comprising all known HIV and SIV clades and sequences as well as derivatives thereby increasing the yields of GagPol production by the use of codon optimized GagPol genes or derivatives thereof as compared to the expression rates achieved from the wildtype gene or wildtype like genes characterized by clustered point mutations - if driven by the same cellular or viral promotor/enhancer unit - as a result of consequent adaptation of the codon usage to that of frequently expressed mammalian genes

[0071] A preferred object of the present invention is the construction of GagPol based vaccines and in particular HIV

9

and SIV based vaccines, especially modified also to improve safety of GagPol derived candidate vaccines if delivered e.g. as a DNA vaccine, by any kind of infectious, replicating or nonreplicating vector or if administered as GagPol derived virus-like particles by consequent and strict codon usage adaptation in complete absence of known cis-acting elements or transactive proteins such as Rev and Tat.

[0072]    Another preferred object of the present invention is the construction of GagPol based vaccines by reducing the opportunities for reconstituting infection competent hybrid retroviral particles by recombination events between nucleic acids encoding the packaging proteins and sequences encoding endogenous retroviruses.

[0073]    Another preferred object of the present invention is the construction of GagPol based vaccines by reducing the risk for incorporating the RNA species encoding Gag, GagPol or derivatives thereof into subviral Gag or GagPol derived particles by more than 25 fold as compared to expression constructs containing the 5'-UTR or encoding the wildtype Gag and GagPol derived genes and by diminishing the risk of packaging the very same RNA into endogenous retroviral particles thereby reducing the risk of recombination between the nucleic acid derived from a vaccine construct encoding Gag, GagPol or derivatives thereof according to the present invention with DNA or RNA derived from an infection by HIV, SIV or any lenti- or retrovirus or otherwise related virus.

[0074]    This results in safe and efficient DNA constructs being useful for the production and delivery of safe GagPol derived vaccine constructs.

[0075]    In order to further increase the safety of the vaccine constructs, codon optimized Gag or GagPol reading frames may be modified at distinct sites.

[0076]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the myristilation of Gag-precursors is inhibited (e.g. position 4 - 6 in SEQ ID No:2 and position 4 - 6 in SEQ ID No:1 are changed to alanin or valin encoding nucleotides), thereby preventing budding of Gag or GagPol derived particles and increasing the amount of endogenously produced antigen.

[0077]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the Gag and Pol coding regions are using the same reading frame by introducing a ribosomal frameshift by either introducing one nucleotide or deleting two nucleotides at any position, without creating a premature stop codon but within the Gag coding region Preferably, between the position of the natural frameshift target site (slippery sequence; corresponding to position 1297 SEQ ID No:1 and position 1294 SEQ ID No:2) and the Gag stop codon (specified by SEQ ID No:3), thereby increasing the amount of synthesized pol gene products.

[0078]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the protease activity is rendered inactive by a deletion comprising the complete or part of the PR gene or by one or several short continuous in frame deletions of various length, Preferably, not exceeding 5 amino acids and more Preferably, comprising a single codon, most Preferably,, encoding the aspartic acid at the PR active site. Also, the PR may be rendered inactive by point mutations at various sites, Preferably, by a single point mutation any position most Preferably, at the active site (e.g. position 1632 - 1635 in SEQ ID No:2 and position 1575 - 1577 in SEQ ID No:1 are changed to asparagin or alanin encoding nucleotides).

[0079]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the active site of the reverse transcriptase gene (RT) is rendered inactive by a deletion comprising the complete or part of the RT gene or one or several short continuous in frame deletions of various length, Preferably, not exceeding 5 amino acids and more Preferably, comprising a single codon encoding an amino acid at any position, most preferable within the active site being critical for RT function. Also, the RT may be rendered inactive by point mutations at various sites, Preferably, by a single point mutation at the active site (e.g. position 2349 - 2351 in SEQ ID No:2 and position 2136 - 2138 in SEQ ID No:1 are changed to asparagin or glutamic acid encoding nucleotides).

[0080]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the active site of the reverse transcriptase gene (RT) is dislocated e.g. at the very C-terminus of the synthetic GagPol gene in order to inhibit polymerase activity thereby preventing the production of potentially hazardous nucleotide sequences and/or reconstitution of infectious or by other means hazardous particles.

[0081]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the other lentiviral genes such as nef, rev, tat, which are rendered biological inactive (e.g. by gene scrambling), are inserted into selected parts of the Gag or GagPol reading frames, Preferably, into the active site of RT in order to broaden immunogenicity and safety.

[0082]    Preferably, the Gag or GagPol based synthetic reading frame is modified in such a way that the potentially hazardous genes, like the integrase (IN) are deleted. Alternatively, the enzyme activity may by knocked out by one deletion or several short continuous in frame deletions of various length, Preferably, not exceeding 5 amino acids and more Preferably, comprising a single codon. Also, the integrase may be rendered inactive by point mutations at various sites, Preferably, by a single point mutation (e.g. position 3930 - 3932 in SEQ ID No:2 and position 3957 - 3959 in SEQ ID No:1 are changed to glutamin encoding nucleotides).

[0083]    Alternatively, the Gag or GagPol based synthetic reading frame is modified by any combination of the above described possibilities. An example of a HIV-1 GagPol derivative modified as described above, is depicted in Fig. 9

10

and specified in SEQ ID No: 4.

[0084] Strict adherence to the codon usage of highly expressed mammalian genes also leads to an increase of the overall GC, GG and CC content and thereby introduces potentially immunomodulatory nucleic acid motifs that increase humoral and cellular immune responses at least 2-fold as compared to the corresponding wildtype gene or wildtype like genes that were rendered partially Rev independent by clustered point mutations.

[0085] The GagPol product or derivatives thereof encoded by the gagpol genes of the present invention may be stably, inducibly or transiently expressed in any primary or lab adapted cell or cell line of human, mammalian or non-mammalian origin in order to produce GagPol antigens, GagPol containing virus-like particles or derivatives thereof.

[0086] The expression of codon optimized GagPol or derivatives thereof in cell culture as well as *in vivo* following immunization with DNA or vector constructs specified below may be driven by any tissue- or celltype-specific cellular promotor/enhancer such as muscle kreatin kinase or MHC class I promotor/enhancer or by any viral (e.g. CMV immediate early; Rous Sarcoma Virus promotor/enhancer, Adenovirus major/late promotor/enhancer) or nonviral e.g. synthetic promotor/enhancer.

[0087] GagPol expression may be either constitutive or inducible e.g. by addition of a specific inducer such as ecdysone in the case of a hormone inducible promotor enhancer or by removal of a repressor (Tet on/off) or repressor gene (Cre/Lox).

[0088] The delivery of codon optimized GagPol genes or derivatives thereof into cells *in vitro* may be mediated by any means of transfecting or introducing a plasmid or plasmid derivative such as "midges" (closed linear DNA) encoding the synthetic GagPol gene or the above variants thereof into cells.

[0089] The delivery of codon optimized GagPol genes or derivatives *in vivo* may be mediated by injection of the plasmid, plasmid derivative or infectious, replicating or nonreplicating vector into any site, Preferably, intramascularly, subcutaneously or intradermally, co-administered with any kind of adjuvant such as liposomes, ISCOMS, alum or via biodegradable particles, either directly or using technical devices like particle gun, biojector or by any other means.

[0090] The delivery of codon optimized GagPol genes or derivatives into cells *in vitro* and *in vivo* may be also mediated by infectious recombinant viral vectors such as recombinant Retroviruses, Vaccinia- or Poxviruses, Adenoviruses, Alphaviruses, Herpesviruses, Baculoviruses or any other recombinant virus, subviral components bridging transfection and infection procedures like e.g. Virosomes comprising nucleic acid/protein aggregates containing e.g. Influenca hemagglutinin, or bacterial vectors such as recombinant Listeriae or Salmonellae or any other type of infectious, replicating or non-replicating vector.

Example 1:

[0091] *Construction of the synthetic gag gene.* All subsequent numbering of HIV-1 wildtype nucleotide sequences correspond to the HIV-1 isolate BH10 (GenBank Accession: M15654). All subsequent numbering of synthetic HIV-1 Gag encoding reading frames correspond to the start codon of the respective coding region. Position of restriction sites are defined by their cleavage site.

[0092] A synthetic sequence coding for the HIV-1$_{IIIB}$ Pr55$^{gag}$ polyprotein was generated by translating the amino-acid sequence of the gag coding region (nucleotides 112 - 1650) into a synthetic-gag (syngag) coding sequence using a codon-usage occurring most frequently in mammalian cells as depicted in Seq ID No: 5. In order to fragmentize the syngag reading-frame, unique restriction sites were generated at positions 5 (*Nar*I), 290 (*Sty*I), 487 (*Stu*I), 696 (*Sac*II), 852 (*Eco*RV), 1107 (*Bst*EII) and 1307 (*Bgl*I) by silent mutations. For cloning, additional restriction sites *Kpn*I (95), *Bam*HI, *Xho*I and *Sac*I were introduced within non-coding regions. Synthetic fragments spanning regions between restriction sites *Kpn*I/*Sty*I (F1), *Sty*I/*Stu*I (F2), *Stu*I/*Eco*RV (F3), *Eco*RV/*Bst*EII (F4), *Bst*EII/*Bgl*I (F5) and *Bgl*I/*Sac*I (F6) were created by stepwise PCR-amplification with overlapping 60 nt long oligonucleotides and subcloned using the pCR-Script™ Amp SK(+) Cloning Kit (Stratagene, Heidelberg, Germany) following the manufactures instructions. The unique restriction sites *Kpn*I, *Sty*I, *Stu*I, *Eco*RV, *Bst*EII, *Sau*I and *Sac*I were used to assemble the fragments into a complete reading frame. Finally, full-length syngag was placed into the *Kpn*I and *Xho*I restriction sites of of pcDNA3.1 (+) expression vector (Invitrogen, Leek, The Netherlands) under the transcriptional control of the immediate-early promoter-enhancer of the Cytomegalovirus (CMV) allowing constitutive transcription in mammalian cells.

[0093] In order to eliminate inhibitory sequences within the HIV-1 gag reading frame, a synthetic gene was constructed, as described above, encoding the entire Pr55$^{gag}$ polyprotein via a codon-usage occurring most frequently in highly expressed mammalian genes. Few deviations from strict adherence to the optimized codon usage were made to accommodate the introduction of unique restriction sites at approximately 250 to 300 nt intervals (Seq ID No: 5). A comparison of synthetic and wildtype Gag coding sequences is shown in Fig. 2 A demonstrating that almost every wobble position within the wildtype coding region was altered by construction of the synthetic gene. The Wisconsin genetics computer group (gcg) software package was used to compare the varying composition of the wildtype and synthetic gag reading frame. Codon optimization resulted in an increased G/C and decreased A/U(T) content, as well as in a 5 fold increase in immune stimulatory CpG motifs (Fig. 2 B).

11

[0094]. Several wildtype expression vectors were established by means of standard molecular biology techniques in order to compare expression of synthetic and wildtype gag reading frames. Subgenomic HIV-1 wildtype sequences were cloned by PCR amplification or restriction digestion of HX10 provirus DNA (Ratner *et al.* 1987, AIDS Res. Hum. Retroviruses. 1987. 3:57-69). The wildtype gag sequence (nucleotides 9-1640) including a 5'-located 103bp long UTR was cloned into the *Kpn*I and *Xho*I restriction sites of pcDNA3.1 (Stratagene) expression vector after PCR amplification of a 1667 nt fragment with primers utr-1 (5'-gcg GGT ACC GAA TTC cga cgc agg act cgg ctt gc-3') and gag-2 (5'-gcc GAG CTC CTC GAG GGA TCC tta ttg tga cga ggg gtc gtt gcc aaa gag-3') resulting in pCMV-UTR-wtgag. Wildtype gag was cloned into the *Kpn*I and *Xho*I restriction sites of pcDNA3.1 (Stratagene) expression vector by PCR amplification of a 1537 nt fragment (103-1640) with primers gag-1 (5'-gcg GGT ACC GAA TTC agg aga gag atg ggt gcg aga gcg tca gta tta agc-3') and gag-2 resulting in pCMV-wtgag. PCR amplification of the UTR with primers utr-1 and utr-2 (5'-gga tG GCG CCc atc tct ctc ctt cta gcc tcc-3') resulted in a 103 nt fragment (9-112) which was cloned into the *Kpn*I and *Nar*I restriction sites of pCMV-syngag, thereby placing the UTR directly 5' of the start-codon of syngag resulting in pCMV-UTR-syngag. *Bgl*II and *BamH*I digestion of HX10 proviral DNA released a 854 nt fragment (6976-7830) containing the RRE and an inefficiently used splice acceptor site (pos. 7734). This RRE containing fragment was cloned into the *BamH*I restriction site of the plasmids pCMV-UTR-wtgag, pCMV-wtgag, and pCMV-UTR-syngag resulting in pCMV-UTR-wtgag-RRE, pCMV-wtgag-RRE and pCMV-UTR-syngag-RRE, respectively. A DNA fragment coding for the constitutive transport element (CTE) RNA element of Simian Mason-Pfizer D-type retrovirus (MPMV, Accession Nr. M12349, nucleotides 7886 to: 8386), was amplified using MPMV proviral DNA as template and primers cte-1 (5' gct aGG ATC Ccc att atc atc gcc tgg aac 3') and cte-2 (5' cga aCT CGA Gca aac aga ggc caa gac atc 3'). The 500 bp fragment was cloned into the *BamH*I and *Xho*I restrction sites of pCMV-wtgag and pCMV-UTR-wtgag resulting in constructs pCMV-wtgag-CTE and pCMV-UTR-wtgag-CTE respectively. A schematic representation of all constructs is summarised in Fig. 1 A.

Example 2:

[0095] *Transfection of Gag encoding expression constructs into mammalian cells.* Cells were transfected by the calcium coprecipitation technique(Graham and Van der EB 1973, Virol. 52, 456-467). The day prior to transfection $2 \times 10^6$ cells were plated on 100 mm-diameter culture dishes and incubated for 24 hours. For transfection, 30µg of Pr55$^{gag}$ expressing plasmids was used and the total amount of DNA adjusted to 45µg with either pcDNA3.1(+) vector or for co-transfection experiments with pCMV-rev. Cells were harvested 48 hours post transfection, washed two times in PBS and then further analyzed. Western blot analysis of cells transfected with Gag encoding expression constructs. Harvested cells were lysed in 0.5% Triton X-100, 100mM Tris/HCl (pH7.4) subjected to repeated freeze/thaw cycles, cleared by centrifugation (20800 x g for 5 min.) and the total amount of protein measured by BIO-RAD Protein-Assay (Bio-Rad Laboratories, Munich, Germany) following the manufactures instructions. 50µg total protein were separated by electrophoresis on a denaturing SDS 12.5% polyacrylamide gel and transferred onto nitrocellulose membrane by electroblotting. Expression of Pr55$^{gag}$ was detected by a HIV-1 p24 specific monoclonal antibody 13-5 (Wolf *et al.* 1990, AIFO 1:24-29) and visualized by chromogenic staining.

[0096] *Gag capture-ELISA of cells transfected with Gag encoding expression constructs.* The p24 capture-ELISA antibodies were kindly provided by Dr. Matthias Niedrig (RKI, Berlin, Germany). 96-well MaxiSorp ELISA plates (Nunc, Wiesbaden, Germany) were coated over night a with 100µl of p24-specific monoclonal antibody 11-G-7 diluted 1:170 with 0.1 M carbonat-buffer (pH 9.5). The plates were washed 6 times with wash-buffer (0.1% Tween 20, 300 mM NaCl, 10 mM $Na_2HPO_4$ $2H_2O$, 1.5 mM $NaH_2PO_4$ $H_2O$). Different amounts of cell-lysats were diluted in 0.5% BSA in wash-buffer added to the wells and incubated over night at 4°C with a second horse-radish-peroxidase conjugated monoclonal antibody (diluted 1:600 in 0.5% BSA in wash-buffer), recognizing a different epitope within p24. After washing the plates 6 times with wash-buffer, antibody conjugates were stained with OPD-solution (Abbott, Wiesbaden, Germany) and absorption OD (495 nm) measured with an ELISA Reader. The concentration of Pr55$^{gag}$ was determined by a calibration curve using different concentrations of purified Pr55$^{gag}$, produced in insect cells using the baculovirus expression system (Wagner *et al.* 1994, Virology 200:162-175).

[0097] As expected, synthesis of Pr55$^{gag}$ from the wildtype gag gene (wtgag) was extremely low in absence of RRE and Rev (Fig. 3 A, lanes 5,6). However, following fusion of RRE downstream of wtgag gene and co-transfection of Rev, Pr55$^{gag}$ production increased only by factor of 1.5-2, suggesting that the Rev/RRE system *per se* was not sufficient to promote substantial Gag expression (Fig. 3 A, lanes 7, 8). In contrast, addition of the authentic 103 bp UTR 5' to the gag gene resulted in a 4-5 fold increase in protein production even in the absence of Rev (Fig. 3 A, lane 1, 3 and 4). Additionally, Rev/RRE interaction led to a further increase in Pr55$^{gag}$ production by a factor of 5-8 (2-4 ng/µg total protein; Fig. 3 A, lane 2) and by several orders of magnitude compared to wtgag-RRE co-transfected with Rev. Substitution of Rev/RRE by CTE-mediated nuclear export depended in very same way on the presence of the authentic UTR resulting on high level Gag expression.. However, CTE mediated Pr55$^{gag}$ production (Fig. 3 B, lane 3) reached only 70% - 80% (1,3 -1,6 ng/µg total protein) of the Rev-dependent Gag expression using Rev/RRE. (Fig. 3 B, lane

1), whereas in absence of UTR or only very little Gag was produced (Fig. 3 B, lanes 4, 5).

[0098] High level expression of Pr55gag was achieved in all cases after transient transfection of the syngag encoding plasmids (Fig. 3 C lane 1-4). Expression levels were not substantially altered by introducing the Rev/RRE system and was neither dependent nor influenced by the presence of UTR (Fig. 3 C lanes 3-4). Pr55gag expression levels exceeded those accomplished by CTE-mediated wtgag expression by more than 130%, and by about 50% to 100% by co-transfection of UTR-wtgag-RRE with Rev (Fig. 3 C, lane 6) and by orders of magnitude to those accomplished by wtgag-RRE whether or not Rev was co-transfected (Fig. 3 C, lane 7, 8). Therefore, codon usage adaptation allows Rev-independent expression in absence of any cis-acting regulatory elements resulting in high yields of Pr55gag production (3.5-6.5 ng/μg total protein).

[0099] These results were confirmed using various cell-lines (Cos7, HeLa) ruling out that cell-type specific factors critically contributed to the observed effects (Fig. 3 D).

Example 3:

[0100] *Immunogenicity of Gag expression vectors after naked DNA vaccination.* Female BALB/c mice (Charles River, Sulzfeld, Germany) were housed under specific pathogen-free conditions and injected at the age of 6 to 12 weeks. In order to evaluate the efficiency and type of immune response induced by either wildtype or synthetic reading frames after DNA vaccination, groups of each 3 mice received a subcutaneous (s.c.) primary injection, at the base of the tail, of a solution containing 100 μg plasmid DNA in a total volume of 100 μl PBS. At week 3 and 6 after the primary injection, mice received a boost injection, respectively.

[0101] Serum was recovered from mice by tail bleed at week two after the boost injection. Antibodies specific for the HIV Pr55gag polyprotein were quantified by an end-point dilution ELISA assay (in duplicate) on samples from individual animals. In brief, a solid phase of PrepCell purified Pr55gag proteins (100 μl of 1 μg/ml per well, overnight at 4° C in a refrigerator) was used to capture Pr55gag-reactive antibodies from the serum (2 h at 37°C), which were then detected with horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG1, IgG2a and total IgG antibodies (1:2000 in PBS, 2% Tween 20, 3% FCS; 100 μl/well; PharMingen, Hamburg, Germany), followed by o-phenlyenediamine dihydrochloride solution (100 μl/well, 20 min at room temperature in the dark; Abbott Laboratories, Abbott Park, IL). End-point titers were defined at the highest serum dilution that resulted in an adsorbance value (OD 492) three times greater than that of the same dilution of a nonimmune serum. The serum of each mouse was assayed and these values were used to calculate the mean and standard deviation for each group of five mice. Single cell suspensions were aseptically prepared from spleens of mice 5 days after the boost immunization. Cells were suspended in α-MEM medium (Gibco) supplemented with 10 mM HEPES buffer, $5 \times 10^{-5}$ β-mercaptoethanole and 10% FCS. 5% of a selected batch of Con A-stimulated rat spleen cell supernatants (Poggi et *al. 1994*, Eur J Immunol. 1994 Sep;24(9):2258-61) were further added to the culture medium as a source of growth factors. Responder cells ($3 \times 10^7$) were cocultured with $1.5 \times 10^6$ syngenic, V3$_{LAI}$ peptide-pulsed P815 cells (irradiated with 20,000 rad) in 10 ml tissue culture medium in upright 25 cm$^3$ tissue culture flasks in a humidified atmosphere with 7% CO$_2$ at 37°C. Cytotoxic effector populations were harvested after 6 days of *in vitro* culture. Serial dilutions of effector cells were cultured with $2 \times 10^3$ target cells in 200 μl round-bottom wells. Targets were autologous A20 cells ($2 \times 10^4$/ml) incubated overnight at 37°C with $10^{-8}$ M of a 18-mer V3$_{LAI}$ peptide or a 23-mer p24CA$_{LAI}$ peptide. Non peptide-pulsed cells were used as a negative control. Target and control A20 cells were labeled with $^{51}$Cr (1 hour at 37°C, 20 μCi/$10^6$ cells) prior to being added to the effector cells. After a 4 hour incubation at 37°C, 100 μl of supernatant were collected for γ-counting. The percentage specific release was calculated as [(experimental release-spontaneous release) / (total release-spontaneous release)] x 100. Total counts were measured after adding 1% Triton X-100 to the labeled target cells. Spontaneously released counts were always less than 20% of the total counts. Data shown are the mean of triplicate cultures. Standard errors of the mean of triplicate data were always less than 20% of the mean.

[0102] As depicted in Fig. 4, synthetic genes derived from lentiviral pathogens are capable of inducing a strong humoral and cellular immune response. Only in mice immunized with synthetic gag plasmids high levels of specific antibodies could be detected. A Th-1 response induces more likely a cellular immune response, which was verified by CTL chromium release assay of freshly prepared spleen cells of the immunized mice. Again, only mice immunized with a Gag expression plasmid encoded with an optimized codon usage was capable of inducing a strong cytolytic activity and therefore cell mediated immune response.

Example 4:

[0103] *Construction of synthetic genes coding for packaging functions of a lentiviral gene transfer vector system.* All subsequent numbering of nucleotide sequences referring to synthetic *GagPol* sequences of HIV-1 or SIV correspond to the start codon of the respective coding region. Position of restriction sites are defined by their cleavage site.

[0104] *Design of a codon optimized HIV-1 GagPol gene encoding packaging functions.* A synthetic sequence coding

for the HIV-1$_{IIIB}$ (BH10, GenBank Accession: M15654) Pr160$^{GagPol}$ polyprotein was generated by backtranslating the amino-acid sequence of the *GagPol* coding region (nucleotides 1 - 4343), except for the region where *gag* and *pol* reading frames are overlapping (nucleotides 1298 - 1539), and by using a codon-usage occurring most frequently in mammalian cells (as depicted in Seq ID No:2). In order to fragmentize this synthetic *GagPol* reading-frame (Hgp$^{syn}$), restriction sites were introduced at positions 290 (*Sty*I), 487 (*Stu*I), 1056 (*Pfl*MI), 1107 (*Bst*EII), 1549 (*Bal*I), 1852 (*Sau*I), 2175 (*Acc*I), 2479 (*Bst*XI), 2682 (*Eco*NI), 2867 (*Sph*I), 3191 (*Pfl*MI), 3370 (*Bcl*I), 3588 (*Bal*I), 3791 (*Xma*III), and 4105 (*Sty*I) by silent mutations. For cloning, additional restriction sites *Kpn*I, *Xho*I and *Sac*I were introduced within non-coding regions. Synthetic fragments spanning regions between restriction sites *Kpn*I/*Sty*I (F1), *Sty*I/*Stu*I (F2), *Stu*I/ *Bst*EII (F3), *Bst*EII/*Bal*I (F4), *Bal*I/*Sau*I (F5), *Sau*I/*Acc*I (F6), *Acc*I/*Bst*XI (F7), *Bst*XI/*Eco*NI (F8), *Eco*NI/*Sph*I (F9), *Sph*I/ *Pfl*MI (F10), *Pfl*MI/*Bcl*I (F11), *Bcl*I/*Bal*I (F12) *Bal*I/*Xma*III (F13), *Xma*III/*Sty*I (F15) and *Sty*I/*Sac*I (F16) were created by a stepwise PCR-amplification of overlapping 60 nt long oligonucleotides and subcloned using the pCR-Script™ Amp SK(+) Cloning Kit (Stratagene, Heidelberg, Germany) following the manufactures instructions. The unique restriction sites *Kpn*I, *Sty*I, *Stu*I, *Bst*EII, and *Sac*I were used to assemble the fragments F1 to F4 into one intermediate fragment (F1-4). The unique restriction sites *Kpn*I, *Bal*I, *Sau*I and *Sac*I were used to assemble the fragments F5 to F7 into a second intermediate fragment (F5-7). The unique restriction sites *Kpn*I, *Bst*XI, *Eco*NI, *Sph*I, *Pfl*MI, *Bcl*I, *Bal*I, *Xma*III, *Sty*I and *Sac*I were used to assemble the fragments F8 to F16 into a third intermediate fragment (F8-16). The intermediate fragments (F1-4, F5-7 and pF8-16) were finally assembled into the full-length synthetic *GagPol* reading frame using restriction sites *Kpn*I/*Bst*EII, *Bst*EII/*Acc*I, and *Acc*I/*Sac*I respectively, and placed into the *Kpn*I and *Sac*I restriction sites of the pCR-Script™ Amp SK(+) cloning vector (Stratagene, Heidelberg, Germany). In order to obtain high level, constitutive expression in mammalian cells the synthetic *GagPol* coding region was placed into the *Kpn*I and *Xho*I restriction sites of pcDNA3.1(+) expression vector (Invitrogen, Leek, The Netherlands) under the transcriptional control of the immediate-early promoter-enhancer of the Cytomegalovirus (CMV) resulting in plasmid Hgp$^{syn}$ (encoding HIV-1 GagPol). *Design of a codon optimized SIV GagPol gene encoding packaging functions.* A synthetic A synthetic sequence coding for the SIV$_{mac239}$ (SIVmm239, GenBank Accession: M33262) Pr160$^{GagPol}$ polyprotein was generated by backtranslating the amino-acid sequence of *the GagPol* coding region (nucleotides 1 - 4358), except the region where *gag* and *pol* reading frames are overlapping (nucleotides 1299 - 1533), and by using a codon-usage occurring most frequently in mammalian cells. In order to fragmentize this synthetic SIV derived *GagPol* reading frame, restriction sites were introduced at positions 120 (*Bst*XI), 387 (*Xma*III), 650 (*Pst*I), 888 (*Nde*I), 1224 (*Apa*I), 1422 (*Nsp*I), 1583 (*Ehe*I), 1899 (*Sau*I), 2019 (*Bsa*BI), 2372 (*Eco*NI), 2651 (*Sac*II), 2785 (*Bcl*I), 3025 (*Bam*HI), 3268 (*Sty*I), 3573 (*Bal*I), 3745 (*Sau*I), 3923 (*Pfl*MI), and 4188 (*Apa*I) by silent mutations. For cloning, additional restriction sites *Kpn*I, *Xho*I and *Sac*I were introduced within non-coding regions. Synthetic fragments spanning regions between restriction sites *Kpn*I/ *Bst*XI (F1), *Bst*XI/*Xma*III (F2), *Xma*III/*Pst*I (F3), *Pst*I/*Nde*I (F4), *Nde*I/*Apa*I (F5), *Apa*I/*Nsp*I (F6), *Nsp*I/*Ehe*I (F7), *Ehe*I/ *Sau*I (F8), *Sau*I/*Bsa*BI (F9), *Bsa*BI/*Eco*NI (F10), *Eco*NI/*Sac*II (F11), *Sac*II/*Bcl*I (F12), *Bcl*I/*Bam*HI (F13), *Bam*HI/*Sty*I (F14), *Sty*I/*Bal*I (F15), *Bal*I/*Sau*I (F16), *Sau*I/*Pfl*MI (F17), *Pfl*MI/*Apa*I (F18) and *Apa*I/*Sac*I (F19) were created by a stepwise PCR-amplification of overlapping 60 nt long oligonucleotides and subcloned using the pCR-Script™ Amp SK (+) Cloning Kit (Stratagene, Heidelberg, Germany) following the manufactures instructions. The unique restriction sites *Kpn*I, *Bst*XI, *Xma*III, *Pst*I, *Nde*I, *Apa*I and *Sac*I were used to assemble the fragments f1 to f6 into one intermediate fragments (f1-6). The fragments f7 to f11 were assembled into a second intermediate fragment (f7-11) using the unique restriction sites *Kpn*I, *Ehe*I, *Sau*I, *Bsa*BI, *Eco*NI and *Sac*I. The unique restriction sites *Kpn*I, *Bcl*I, *Bam*HI, *Sty*I, *Bal*I, *Sau*I, *Pfl*MI, *Apa*I and *Sac*I were used to assemble the fragments f12 to f19 into a third intermediate fragment (f12-19). The intermediate fragments f1-6, f7-11 and f12-19 were finally assembled into the full-length synthetic *GagPol* gene of SIV using restriction sites *Kpn*I/*Nsp*I, *Nsp*I/*Sac*II, and *Sac*II/*Sac*I respectively, and placed into the *Kpn*I and *Sac*I restriction sites of the pCR-Script™ Amp SK(+) cloning vector (Stratagene, Heidelberg, Germany). In order to obtain high level, constitutive expression in mammalian cells the SIV derived synthetic *GagPol* gene was placed into the *Kpn*I and *Xho*I restriction sites of pcDNA3.1(+) expression vector (Invitrogen, Leek, The Netherlands) under the transcriptional control of the immediate-early promoter-enhancer of the Cytomegalovirus (CMV) resulting in plasmid Sgp$^{syn}$ (encoding SIV$_{mac239}$ GagPol).

[0105]  Codon usage of the 4.3 kb long *GagPol* genes of HIV-1 and SIV was optimized for expression in human cells by assembling the entire reading frame from synthetic oligonucleotides. The synthetic genes encode GagPol proteins which have the same amino acid sequence as wildtype SIV or HIV-1 GagPol. On the nucleotide level there is an overall identity of 69,27 % for the synthetic and wildtype *GagPol* of HIV-1 and of 73,24% for the synthetic and wildtype SIV *GagPol* genes (Table 1).

Table 1:

| Genetic distances between the different SIV and HIV-1 GagPol coding sequences | | | | |
|---|---|---|---|---|
| GagPol gene | SIVsyn | SIVwt | HIVsyn | HIVwt |
| SIVsyn | 0.00 | 29.81 | 36.16 | 72.21 |
| SIVwt | - | 0.00 | 69.78 | 54.72 |
| HIVsyn | - | | 0.00 | 30.73 |
| HIVwt | - | - | - | 0.00 |

[0106] Distance matrix of SIV$_{mac239}$ and HIV$_{IIIB}$ derived synthetic (syn) and wildtype (wt) GagPol coding sequences. Numbers indicate the percentage of missmatches between the compared GagPol sequences.

[0107] Optimizing the codon usage of HIV-1 GagPol reduces the homology to wildtype SIV GagPol by 41.48% and vice versa (Table 1). The longest stretch of nucleotide identity between synthetic and wildtype GagPol is 357 bp in SIV and 306 bp in HIV-1. This stretch is located in the p69$^{gag}$/p6$^{pol}$ overlap region containing the slippery sequence and a stem loop structure, both of which are required for proper ribosomal frameshifting.

Example 5:

[0108] *Expression of synthetic GagPol genes.* Rev-independent expression of the synthetic HIV-1 and SIV GagPol genes was assayed in transient transfection experiments followed by characterisation of the GagPol particles by conventional immunoblot analysis. Briefly, the day prior to transfection 2x10$^6$ H1299 cells were plated on 100 mm-diameter culture dishes and incubated for 24 hours. Cells were transfected using 45µg of the synthetic GagPol expressing plasmids or proviral DNA (HX10 and SIV$_{mac239}$) with the calcium phosphate coprecipitation method as described (Graham and Van der EB 1973, Virol. 52, 456-467). Cell culture supernatants were harvested 72 hours post transfection. To block processing of GagPol, plated-cells were cultured and transfected in the presence of 10 µM Saquinavir. GagPol particles were purified by separating the culture supernatants on 10-60% sucrose gradients as previously described (Wagner et al. 1994, Virology 200:162-175). GagPol particles banded at the expected density of 1.13 to 1.19 g/cm$^2$, typical for HIV virions. The fractions with highest particle content were dialyzed against PBS and further separated by electrophoresis on a denaturing SDS 12.5% polyacrylamide gel prior to transfere onto nitrocellulose membrane by electroblotting. HIV-1 p24 capsid (CA) and SIV p27 CA and their precursor proteins were detected by the monoclonal antibodies 13-5 (Wolf et al. 1990, AIFO 1:24-29.) and 55-2F12 (NIH, Rockville, USA), respectively.

[0109] Western Blot analyses from supernatants of cells transfected with expression plasmids for the synthetic HIV-1 and SIV GagPol genes (Hgp$^{syn}$, Sgp$^{syn}$ in Fig. 5 B) revealed efficient expression of GagPol in the absence of Rev (Fig. 5 A, B). The GagPol levels obtained with the synthetic genes were significantly higher as the levels obtained by transfection of infectious proviral constructs in the presence of Rev as assessed by a commercial capture ELISA format. Subviral particles were readily released from cells transfected with the synthetic GagPol expression plasmids and sedimentate at a density of about 1,16 g/ml, comparable to wildtype SIV or HIV-1 virions. No differences were observed in the processing of the GagPol precursor proteins encoded by the wildtype viruses and the corresponding synthetic GagPol genes (Fig. 5 A, B).

Example 6:

[0110] *Vector production and infection of target cells.* The 293T cells were transfected with the calcium phosphate coprecipitation method as described (Graham and Van der EB 1973, Virol. 52, 456-467). The plasmids pHIT60 (MLV GagPol expression plasmid), pHIT-G (VSV-G expression plasmid), SgpΔ2 (SIV-GagPol expression plasmid also expressing vif, vpr, vpx, tat and rev), and ViGΔBH (self-inactivating SIV vector expressing the green fluorescent protein gene (GFP) under the control of the promoter and enhancer from spleen focus forming virus) have been described previously (Graham and Van der EB 1973, Virol. 52, 456-467; Fouchier et al. 1997, EMBO 16, 4531-4539). The Murine leukemia virus (MLV) vector pLEGFP-N1, which contains the GFP gene under the control of an internal CMV promoter, was obtained from Clontech. The supernatant of the transfected cells was passed through a 0.45 µm filter and stored in aliquots at -80°C. The vector stocks were assayed for CA antigen levels using an HIV p24 enzyme-linked immunosorbent assay (ELISA) kit (Innogenetics, Gent, Belgium). The standard included in the ELISA kit was used for quantification of HIV-1 p24 CA antigen levels. Recombinant SIV p28 CA antigen from the SIV p28 antigen capture assay from the AIDS vaccine program (NCI-Frederic cancer research and development center) served as a standard for the quantification of SIV p28 levels.

[0111]. To determine vector titers, 293 cells were seeded at 2.5 x 10$^4$ cells per well of a 24-well plate. One day later, the medium was removed and cells were incubated for 2 to 4 hours with serial dilutions of the vector preparations in a total volume of 200 μl. Fresh medium was added and the number of GFP positive cells was determined two days after infection. To arrest 293 cells in the GI/S phase of the cell cycle, cells were seeded in a 96-well plate at a density of 2 to 5 x 10$^4$ cells per well at a concentration of 1, 3, or 5 μg aphidicolin/ml (Sigma). The titer of the vector stocks was determined on these cells as described above with aphidicolin being present during the entire culture period.

[0112]. The functional integrity of the GagPol proteins being encoded by synthetic genes was analysed by cotransfection of the synthetic GagPol expression plasmids and an expression plasmid for VSV-G (pHIT-G) with the SIV vector ViGΔBH (Fig. 6). This is a self-inactivating vector, which expresses the green fluorescent protein gene (GFP) under the control of an internal promoter. The vector titers in the supernatant of the transfected cells were in the range of 1x10$^6$ GFP-forming units/ml supernatant (Table 2). Similar titers were obtained after cotransfection of ViGΔBH with a wildtype GagPol expression plasmid of SIV (SgpΔ2). The synthetic HIV-1 GagPol expression plasmid also allowed efficient transfer of the SIV vector (Table 2). The infectivity of the vector particles varied less than 2-fold between the different GagPol expression plasmids used (Table 2). Therefore, HIV-1 GagPol must recognise all cis-acting sequences of SIV required for packaging, reverse transcription, and integration with similar efficiency as SIV. As using packaging functions the results clearly extent a previous report on the packaging of HIV-2 RNA, which is closely related to SIVmac, in HIV-1 particles (Kaye and Lever, 1998, J.Virol. 72, 5877-5885).

Table 2.

| Titer and infectivity of an SIV vector packaged by synthetic SIV and HIV-1 GagPol expression plasmids | | | |
|---|---|---|---|
| | Titer[a] | | |
| Plasmids transfected | Experiment 1 | Experiment 2 | Infectivity[b] |
| SgpΔ2, ViGΔBH, VSV-G | 5.0x10$^6$ | 1.4 x10$^6$ | 5.9x10$^5$ |
| Sgp$^{syn}$, ViGΔBH, VSV-G | 1.0x10$^6$ | 2.7x10$^5$ | 3.4x10$^5$ |
| Hgp$^{syn}$, ViGΔBH, VSV-G | 2.0x10$^6$ | 2.0x10$^6$ | 3.1x10$^5$ |
| ViGΔBH, VSV-G | <5 | <5 | n.a. |

[a] GFP forming units (GFU) per ml supernatant of 293T cells transfected with the indicated plasmids;

[b] GFU/ng CA antigen in the supernatant of transfected 293T cells; n.a.: not applicable.

[0113] Omitting the GagPol expression plasmid reduced the titer below the level of detection (Table 2), which argues against VSV-G mediated pseudotransduction. In addition, the percentage of GFP positive cells transduced with ViGΔ-BH, which had been packaged by Sgp$^{syn}$, Hgp$^{syn}$, or SgpΔ2, did not decrease significantly during a six week observation period (data not shown), which suggests integration of the vector into the host genome. As previously observed for another SIV vector packaged by SgpΔ2 and pHIT-G replication competent recombinants could not be detected during an 8 week observation period in ViGΔBH vector preparations packaged by SgpΔ2, Sgp$^{syn}$, or Hgp$^{syn}$ and pHIT-G (data not shown). The transduction efficiency of vectors produced with the synthetic GagPol genes for non-dividing cells was assessed by arresting the target cells in the G1 phase of the cell cycle by aphidicolin treatment. The titer of an MLV-based vector in growth-arrested cells was reduced to background levels (Fig. 7). In contrast, the SIV vector titer was only slightly reduced by aphidicolin treatment. Non-dividing cells could be transduced with the SIV vector independent of the GagPol expression plasmid used for the production of the vector particle (Fig. 7).

Example 7:

[0114] CEMx174 cells (5 x 10$^5$) were infected for 4 hours with 0.5 ml of the vector preparation. The cells were washed three times with PBS and subsequently cultured in 5 ml medium. Infected cultures were split 1:5 to 1:10 twice weekly for eight weeks. The capsid antigen levels in theses cultures were determined 1 and 4 days, and 2, 4, 6 and 8 weeks after infection using the HIV-Innogenetics ELISA and the recombinant SIV p28 antigen as a standard. To determine the emergence of RCR on CEMx174-SIV-SEAP cells (Means et al., 1997, J.Virol. 71, 7895-7902), 1x10$^5$ cells were infected with 1 ml supernatant of the transfected cells for two hours and washed once in PBS. Infected cultures were cultured in 5 ml medium and split 1:5 to 1:10 twice weekly for four weeks. Secreted alkaline phosphatase activity in the supernatant of these cultures was determined at different time points after infection with the Phospha-Light-kit (Tropix, Bedford, MA, USA) as described by the manufacturer. The titer of RCRs was determined by limiting dilution. CEMx174-SIV-SEAP cells (2x10$^5$) were incubated with 1 ml supernatant in a final volume of 2 ml. After two hours, six 200 μl aliquots were transferred to a 96-well plate and used as a starting point for 4-fold serial dilutions. Cultures were split 1:10 after one week. Syncytia formation was scored 1 and 2 weeks after infection and the TCID$_{50}$ of the RCRs

16

was calculated as described (Johnson and Byington, 1990, Quantitative assays for virus infectivity. In Techniques in HIV research. A. Aldovini and B.D. Walker, eds. (New York: Stockton Press), pp. 71-76. 1990).

[0115]   *Detection of replication-competent recombinants.* The potential emergence of replication competent recombinants (RCR) from our vector preparation was analysed in CEMx174 cells, which support replication of SIV and SIV hybrid viruses utilising a heterologous *env* gene (Reiprich et al., 1997, J.Virol. *71*, 3328-3331). These cells were infected with 0.5 ml of ViGΔBH vector preparations (for titers see example 6, Table 2) packaged by SgpΔ2, Sgp[syn], or Hgp[syn] and pHIT-G. Slightly elevated capsid antigen levels (<400 pg SIVp27CA/ml) could be detected one and four days after infection. This has been observed previously and is probably due to carry over CA-antigen from the vector preparation (Schnell et al., 2000, Hum.Gene Ther. *11*, 439-447). More importantly, CA antigen levels were below the level of detection of 20 pg SIVp27/ml two to eight weeks after infection in all cultures and no cytopathic effects could be observed. Therefore, the titer of a potential RCR is below 2 infectious unit/ml. Since no RCR could be detected with the wild-type SIV-*gag-pol* expression plasmid, SgpΔ2, this assay did not reveal potential advantages of the synthetic *gag-pol* expression plasmids. We therefore tested the frequency of emergence of RCR in an assay system, that would only require two homologous recombination events. SgpΔ2, Sgp[syn], or Hgp[syn] was cotransfected with either of two SIV vectors, that still contained env, vif, vpr, vpx, tat, and rev in addition to the GFP reporter gene, but which contained large deletions in *gag-pol*. The CEMx174-SIV-SEAP cells, which release SEAP after infection with SIV due to Tat transactivation (Means et al., 1997, J.Virol. 71, 7895-7902), were infected with 1 ml of the supernatant of the transfected cells. Three days after infection elevated SEAP activity was observed in cultures infected with the SIV vectors packaged by the synthetic gag-pol expression plasmids (Fig. 8). This should be due to transcomplementation of *gag-pol* and subsequent transfer of the SIV vector. However, the SEAP activity returned to background levels during the next 3 weeks (Fig. 8) indicating the absence of RCR. In contrast, increasing SEAP activity three to ten days after infection revealed rapid emergence of RCR in SIV vectors packaged by SgpΔ2. This was accompanied by extensive syncytia formation and cell death. Using a limiting dilution approach, the titer of the RCR in the supernatant of cells transfected with SgpΔ2 and SIV-GFPΔDP or SgpΔ2 and SIV-GFPΔBP was determined to be 130 and 75 TCID$_{50}$/ml, respectively. This demonstrates that if RCR emerge at all with the synthetic *gag-pol* genes, the frequency is at least reduced by a factor of approximately 100 in comparison to a wild-type *gag-pol* expression plasmid.

SEQUENCE LISTING

<110> Wagner, Ralf

<120> Synthetic gagpol genes and their uses

<130> WAG-002

<140> xx
<141> 2000-05-18

<160> 5

<170> PatentIn Ver. 2.1

<210> 1
<211> 4358
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: sequence with
      optimized codons

<400> 1
```
atgggcgtga ggaacagcgt gctgagcggc aagaaggccg acgagctgga gaagatcagg 60
ctgaggccca acggcaagaa gaagtatatg ctgaagcacg tggtgtgggc cgccaacgag 120
ctggacaggt tcggcctggc cgagagcctg ctggagaaca aggagggctg ccagaagatc 180
ctgagcgtgc tggcccccct ggtgcccacc ggcagcgaga acctgaagag cctgtacaac 240
accgtgtgcg tgatctggtg catccacgcc gaggagaagg tgaagcacac cgaggaggcc 300
aagcagatcg tgcagaggca cctggtggtg gagaccggca ccaccgagac catgcccaag 360
accagcaggc ccaccgcccc cagctccggc cgcggcggca actacccgt gcagcagatc 420
ggcggcaact acgtgcacct gccccctgagc cccaggaccc tgaacgcctg ggtgaagctg 480
atcgaggaga agaagttcgg cgccgaggtg gtgcccggct tccaggccct gagcgagggc 540
tgcacccctt acgacatcaa ccagatgctg aactgcgtgg gcgaccacca ggcgccatg 600
cagatcatca gggacatcat caacgaggag gccgccgact gggacctgca gcaccctcag 660
cccgcccctc agcagggcca gctgagggag cccagcggca gcgacatcgc cggcaccaca 720
agcagcgtgg acgagcagat ccagtggatg tacaggcagc agaaccctat ccccgtgggc 780
aacatctaca ggaggtggat ccagctgggc ctccagaagt gcgtgaggat gtacaacccc 840
acaaacatcc tggacgtgaa gcagggacca aaggagccct tccagtcata tgtggacagg 900
ttctacaaga gcctgagggc cgagcagacc gacgccgccg tgaagaactg gatgacccag 960
accctgctga tccagaacgc caaccccgac tgcaagctgg tgctgaaggg cctgggcgtg 1020
aaccccaccc tggaggagat gctgaccgcc tgccagggcg tgggcggccc cggccagaag 1080
```

```
gctaggctga tggccgaggc tctgaaggag gccctggccc ccgtgcccat cccottcgcc 1140
gccgcccagc agaggggacc caggaagccc atcaagtgct ggaactgcgg caaggagggc 1200
cacagcgcca ggcagtgcag ggcccccagg aggcagggct gctggaagtg cggcaagatg 1260
gaccacgtga tggccaagtg ccccgacagg caggccggtt ttttaggcct tggtccatgg 1320
ggaaagaagc cccgcaattt ccccatggct caagtgcatc aggggctgat gccaactgct 1380
ccoccagagg acccagctgt ggatctgcta aagaactaca tgcagttggg caagcagcag 1440
agagaaaagc agagagaaag cagagagaag ccttacaagg aggtgacaga ggatttgctg 1500
cacctcaatt ctctctttgg aggagaccag tagtgaccgc ccacatcgag ggccagcccg 1560
tggaggtgct gctggacacc ggcgccgacg acagcatcgt gaccggcatc gagctgggac 1620
cccactacac ccccaagatc gtgggcggca tcggcggctt catcaacaca aaggagtaca 1680
agaacgtgga gatcgaggtg ctgggcaaga ggatcaaggg caccatcatg accggcgaca 1740
cccccatcaa catcttcggc aggaacctgc tgaccgccct gggcatgagc ctgaacttcc 1800
ccatcgccaa ggtggagccc gtgaaggtgg ccctgaagcc cggcaaggac ggccccaagc 1860
tgaagcagtg gcctctgagc aaggagaaga tcgtggccct gagggaaatc tgcgagaaga 1920
tggagaagga cggccagctg gaggaggccc ctcccaccaa cccctacaac accccacct 1980
tcgccatcaa gaagaaggac aagaacaagt ggaggatgct gatcgacttc agggagctga 2040
acaggggtgac acaggacttc accgaggtgc agctgggcat ccctcacccc gccggcctgg 2100
ccaagaggaa gaggatcacc gtgctggaca tcggcgacgc ctacttcagc atccctctgg 2160
acgaggagtt caggcagtac accgccttca ccctgcccag cgtgaacaac gccgagcccg 2220
gcaagaggta catctacaag gtgctgcccc agggctggaa gggcagcccc gccatcttcc 2280
agtacaccat gaggcacgtg ctggagccct tcaggaaggc caaccccgac gtgaccctgg 2340
tgcagtacat ggacgacatc ctgatcgcct ccgacaggac cgacctggag cacgacaggg 2400
tggtgctcca gagcaaggag ctgctgaaca gcatcggctt cagcacccc gaggagaagt 2460
tccagaagga ccctcccttc cagtggatgg gctacgagct gtggcccacc aagtggaagc 2520
tccagaagat cgagctgccc cagagggaga cctggaccgt gaacgacatc cagaagctgg 2580
tgggcgtgct gaactgggcc gcccagattt accccggcat caagaccaag cacctgtgca 2640
ggctgatccg cggcaagatg acactgaccg aggaggtgca gtggaccgag atggccgagg 2700
ccgagtacga ggagaacaag atcattctga gccaggagca ggagggctgc tactaccagg 2760
agggcaagcc cctggaggcc accgtgatca gagccagga caaccagtgg agctacaaga 2820
tccaccagga ggacaagatc ctgaaggtgg gcaagttcgc caagatcaag aacacccaca 2880
ccaacggcgt gaggctgctg gcccacgtga tccagaagat cggcaaggag gccatcgtga 2940
tctgggocca ggtgcccaag ttccacctgc ccgtggagaa ggacgtgtgg gagcagtggt 3000
ggaccgacta ctggcaggtg acatggatcc ccgagtggga cttcatcagc accctcctc 3060
tggtgaggct ggtgttcaat ctggtgaagg accccatcga gggcgaggag acctactaca 3120
ccgacggcag ctgcaacaag cagagcaagg agggcaaggc cggctacatc accgacaggg 3180
gcaaggacaa ggtgaaggtg ctggagcaga ccaccaacca gcaggccgag ctggaggcct 3240
tcctgatggc cctgaccgac agcggcccca aggccaacat catcgtggac agccagtatg 3300
tgatgggcat catcaccggc tgccccaccg agagcgagag caggctggtg aaccagatca 3360
tcgaggagat gattaagaag agcgagattt acgtggcctg ggtgcccgcc cacaagggca 3420
tcggcggcaa ccaggagatc gaccacctgg tgagccaggg catcaggcag gtgctgttcc 3480
tggagaagat cgagcccgcc caggaggagc acgacaagta ccacagcaac gtgaaggagc 3540
tggtgttcaa gttcggcctg cccaggatcg tggccaggca gatcgtggac acctgcgaca 3600
agtgccacca gaagggcgag gccatccacg ccaggccaa cagcgacctg ggcacctggc 3660
```

19

```
agatggactg cacccacctg gagggcaaga tcatcatcgt ggccgtgcac gtggctagcg 3720
gcttcatcga ggccgaggtg atccctcagg agaccggcag gcagaccgcc ctgttcctgc 3780
tgaagctggc cggcaggtgg cccatcaccc acctgcacac cgacaacggc gccaacttcg 3840
ccagccagga ggtgaagatg gtggcctggt gggccggcat cgagcacacc ttcggcgtgc 3900
cctacaaccc ccagagccag ggcgtggtgg aggccatgaa ccaccacctg aagaaccaga 3960
tcgacaggat cagggagcag gccaacagcg tggagaccat cgtgctgatg gccgtgcact 4020
gcatgaactt caagaggagg ggcggcatcg cgacatgac ccccgccgag aggctgatca 4080
acatgattac caccgagcag gagatccagt ccagcagag caagaacagc aagttcaaga 4140
acttcagggt gtattacagg gagggcaggg accagctgtg gaagggcccc ggcgagctgc 4200
tgtggaaggg cgaggcgct gtgatcctga aggtgggcac cgacatcaag gtggtgccca 4260
ggaggaaggc caagatcatc aaggactacg cggcggcaa ggaggtggac agcagcagcc 4320
acatggagga caccggcgag gccagggagg tggcctga            4358
```

```
<210> 2
<211> 4343
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: sequence with
      optimized codons

<400> 2
atgggcgcca gggccagcgt gctgagcggc ggcgagctgg acaggtggga gaagatcagg 60
ctgaggcccg gcggcaagaa gaagtataag ctgaagcaca tcgtgtgggc cagcagggag 120
ctggagaggt tcgccgtgaa ccccggcctg ctggagacca gcgagggctg caggcagatc 180
ctgggccagc tgcagcccag cctgcagacc ggcagcgagg agctgaggag cctgtacaac 240
accgtggcca ccctgtactg cgtgcaccag aggatcgaga tcaaggacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tccaagaaga aggcccagca ggccgccgcc 360
gacaccggcc acagcagcca ggtgagccag aactacccca tcgtgcagaa catccagggc 420
cagatggtgc accaggccat cagccccagg accctgaacg cctgggtgaa ggtggtggag 480
gagaaggcct tcagccccga ggtgatcccc atgttcagcg ccctgagcga gggagccacc 540
ccccaggacc tgaacaccat gctgaacacc gtgggcggcc accaggccgc catgcagatg 600
ctgaaggaga ccatcaacga ggaggccgcc gagtgggaca gggtgcaccc cgtgcacgcc 660
ggccccatcg cccccggcca gatgagggag ccccgcggca gcgacatcgc cggcaccacc 720
agcaccctgc aggagcagat cggctggatg accaacaacc cccccatccc cgtgggcgaa 780
atctacaaga ggtggatcat cctgggcctg aacaagatcg tgaggatgta cagccccacc 840
agcatcctgg atatcaggca gggccccaaa gagcccttca gggactacgt ggacaggttc 900
tacaagaccc tgcgcgccga gcaggccagc caggaggtga agaactggat gaccgagacc 960
ctgctggtgc agaacgccaa ccccgactgc aagaccatcc tgaaggccct gggacccgcc 1020
gccaccctgg aggagatgat gaccgcctgc cagggcgtgg gcggccccgg ccacaaggcc 1080
agggtgctgg ccgaggccat gagccaggtg accaacaccg ccaccatcat gatgcagagg 1140
```

```
ggcaacttca ggaaccagag gaagatggtg aagtgcttca actgcggcaa ggagggccac 1200
accgccagga actgccgcgc ccccaggaag aagggctgct ggaagtgcgg caaggagggc 1260
caccagatga aggactgcac cgagaggcag gctaattttt tagggaagat ctggccttcc 1320
tacaagggaa ggccagggaa ttttcttcag agcagaccag agccaacagc cccaccattt 1380
cttcagagca gaccagagcc aacagcccca ccagaagaga gcttcaggtc tggggtagag 1440
acaacaactc cccctcagaa gcaggagccg atagacaagg aactgtatcc tttaacttcc 1500
ctcagatcac tctttggcaa cgacccctcg tcacaataaa gatcggtggc cagctgaagg 1560
aggccctgct ggacaccggc gccgacgaca ccgtgctgga ggagatgagc ctgcccggca 1620
ggtggaagcc caagatgatc ggcggcatcg gcggcttcat caaggtgagg cagtacgacc 1680
agatcctgat cgagatctgc ggccacaagg ccatcggcac cgtgctggtg ggccccaccc 1740
ccgtgaacat catcggcagg aacctgctga cccagatcgg ctgcaccctg aacttcccca 1800
tcagccccat cgagaccgtg cccgtgaagc tgaagcccgg catggacggc cctaaggtga 1860
agcagtggcc cctgaccgag gagaagatca aggccctggt ggagatctgc accgagatgg 1920
agaaggaggg caagatcagc aagatcggcc ccgagaaccc ctacaacacc cccgtgttcg 1980
ccatcaagaa gaaggacagc accaagtgga ggaagctggt ggacttcagg gagctgaaca 2040
agaggaccca ggacttctgg gaggtgcagc tgggcatccc ccacccccgcc ggcctgaaga 2100
agaagaagag cgtgaccgtg ctggacgtgg gcgacgccta cttcagcgtg cccctggacg 2160
aggacttcag gaagtatacc gccttcacca tccccagcat caacaacgag accccccggca 2220
tccgctacca gtacaacgtg ctgccccagg gctggaaggg cagccccgcc atcttccaga 2280
gcagcatgac aaagatcctg gagcccttca gaagcagaa ccccgacatc gtgatctatc 2340
agtacatgga cgacctgtac gtgggcagcg acctggagat cggccagcac aggaccaaga 2400
tcgaggagct gaggcagcac ctgctgaggt ggggcctgac cacccccgac aagaagcacc 2460
agaaggagcc cccattcctg tggatgggct acgagctgca ccccgacaag tggaccgtgc 2520
agcccatcgt gctgcccgag aaggacagct ggaccgtgaa cgacattcag aagctggtgg 2580
gcaagctgaa ctgggccagc cagatctacc ccggcatcaa ggtgaggcag ctgtgcaagc 2640
tgctgagggg cacaaaggct ctgaccgagg tgatccccct gaccgaggag gccgagctgg 2700
agctggccga gaacagggag atcctgaagg agcccgtgca cggcgtgtac tacgacccca 2760
gcaaggacct gatcgccgag atccagaagc agggccaggg ccagtggacc taccagatct 2820
accaggagcc cttcaagaac ctgaagaccg gcaagtacgc ccgcatgcgc ggcgcccaca 2880
ccaacgacgt gaagcagctg accgaggccg tgcagaagat caccaccgag agcatcgtga 2940
tctggggcaa gaccccaag ttcaagctgc ccatccagaa ggagacctgg gagacctggt 3000
ggaccgagta ctggcaggcc acctggattc ccgagtggga gttcgtgaac acccctcccc 3060
tggtgaagct gtggtatcag ctggagaagg agcccatcgt gggcgccgag accttctacg 3120
tggacggcgc cgccaacagg gagaccaagc tgggcaaggc cggctacgtg accaacaagg 3180
gccgccagaa ggtggtgccc ctgaccaaca ccaccaacca gaagaccgag ctgcaggcta 3240
tctacctggc cctgcaggac tcaggcctgg aggtgaacat cgtgaccgac agccagtacg 3300
ccctgggcat catccaggcc cagcccgaca gagcgagag cgagctggtg aaccagatca 3360
tcgagcagct gatcaagaag gagaaggtgt acctggcctg ggtgcccgcc cacaagggca 3420
tcggcggcaa cgagcaggtg gacaagctgg tgagcgccgg catcaggaag atcctgttcc 3480
tggacggcat cgacaaggcc caggacgagc acgagaagta ccacagcaac tggagggcta 3540
tggctagcga cttcaacctg cctcccgtgg tggctaagga gatcgtggcc agctgcgaca 3600
agtgccagct gaagggcgag gccatgcacg gccaggtgga ctgcagcccc ggcatctggc 3660
agctggactg cacccacctg gagggcaagg tgatcctggt ggccgtgcac gtggcctccg 3720
```

```
gctacatcga ggccgaggtg atccccgccg agaccggcca ggagaccgcc tacttcctgc 3780
tgaagctggc cggccgctgg cccgtgaaga ccatccacac cgacaacggc agcaacttca 3840
ccagcgccac cgtgaaggcc gcctgctggt gggccggcat caagcaggag ttcggcatcc 3900
cctacaaccc ccagtctcag ggcgtggtgg agagcatgaa caaggagctg aagaagatca 3960
tcggccaggt gagggaccag gccgagcacc tgaagaccgc cgtgcagatg gccgtgttca 4020
tccacaactt caagaggaag ggcggcatcg gcggctacag cgccggcgag aggatcgtgg 4080
acatcatcgc caccgacatc cagaccaagg agctgcagaa gcagatcacc aagatccaga 4140
acttcagggt gtactacagg gacagcagga accctctgtg gaagggcccc gccaagctgc 4200
tgtggaaggg cgagggcgcc gtggtgatcc aggacaacag cgacatcaag gtggtgccca 4260
ggaggaaggc caagatcatc agggactacg gcaagcagat ggccggcgac gactgcgtgg 4320
cctccaggca ggacgaggac tga                                         4343
```

<210> 3

<211> 4341

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: sequence with
     optimized codons

<400> 3
```
atgggcgcca gggccagcgt gctgagcggc ggcgagctgg acaggtggga gaagatcagg 60
ctgagccccg gcggcaagaa gaagtataag ctgaagcaca tcgtgtgggc cagcagggag 120
ctggagaggt tcgccgtgaa ccccggcctg ctggagacca gcgagggctg caggcagatc 180
ctgggccagc tgcagcccag cctgcagacc ggcagcgagg agctgaggag cctgtacaac 240
accgtggcca ccctgtactg cgtgcaccag aggatcgaga tcaaggacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tccaagaaga aggcccagca ggccgccgcc 360
gacaccggcc acagcagcca ggtgagccag aactacccca tcgtgcagaa catccaggc 420
cagatggtgc accaggccat cagccccagg accctgaacg cctgggtgaa ggtggtggag 480
gagaaggcct tcagccccga ggtgatcccc atgttcagcg ccctgagcga gggagccacc 540
ccccaggacc tgaacaccat gctgaacacc gtgggcggcc accaggccgc catgcagatg 600
ctgaaggaga ccatcaacga ggaggccgcc gagtgggaca gggtgcaccc cgtgcacgcc 660
ggccccatcg cccccggcca gatgagggag ccccgcggca gcgacatcgc cggcaccacc 720
agcaccctgc aggagcagat cggctggatg accaacaacc ccccatcccc cgtgggcgaa 780
atctacaaga ggtggatcat cctgggcctg aacaagatcg tgaggatgta cagccccacc 840
agcatcctgg atatcaggca gggccccaaa gagcccttca gggactacgt ggacaggttc 900
tacaagaccc tgcgcgccga gcaggccagc caggaggtga gaactggat gaccgagacc 960
ctgctggtgc agaacgccaa ccccgactgc aagaccatcc tgaaggccct gggaccccgcc 1020
gccaccctgg aggagatgat gaccgcctgc caggggcgtgg gcggcccccgg ccacaaggcc 1080
agggtgctgg ccgaggccat gagccaggtg accaacaccg ccaccatcat gatgcagagg 1140
ggcaacttca ggaaccagag gaagatggtg aagtgcttca actgcggcaa ggagggccac 1200
```

```
accgccagga actgccgcgc ccccaggaag aagggctgct ggaagtgcgg caaggagggc 1260
caccagatga aggactgcac cgagaggcag gctaatttta gggaagatct ggccttccta 1320
caagggaagg ccagggaatt ttcttcagag cagaccagag ccaacagccc caccatttct 1380
tcagagcaga ccagagccaa cagccccacc agaagagagc ttcaggtctg gggtagagac 1440
aacaactccc cctcagaagc aggagccgat agacaaggaa ctgtatcctt taacttccct 1500
cagatcactc tttggcaacg acccctcgtc acaataaaga tcggtggcca gctgaaggag 1560
gccctgctgg acaccggcgc cgacgacacc gtgctggagg agatgagcct gcccggcagg 1620
tggaagccca agatgatcgg cggcatcggc ggcttcatca aggtgaggca gtacgaccag 1680
atcctgatcg agatctgcgg ccacaaggcc atcggcaccg tgctggtggg ccccacccccc 1740
gtgaacatca tcggcaggaa cctgctgacc cagatcggct gcaccctgaa cttccccatc 1800
agccccatcg agaccgtgcc cgtgaagctg aagcccggca tggacggccc taaggtgaag 1860
cagtggcccc tgaccgagga gaagatcaag gccctggtgg agatctgcac cgagatggag 1920
aaggagggca agatcagcaa gatcggcccc gagaacccct acaacacccc cgtgttcgcc 1980
atcaagaaga aggacagcac caagtggagg aagctggtgg acttcaggga gctgaacaag 2040
aggacccagg acttctggga ggtgcagctg ggcatccccc accccgccgg cctgaagaag 2100
aagaagagcg tgaccgtgct ggacgtgggc gacgcctact tcagcgtgcc cctggacgag 2160
gacttcagga agtataccgc cttcaccatc cccagcatca caacgagac ccccggcatc 2220
cgctaccagt acaacgtgct gccccagggc tggaagggca gccccgccat cttccagagc 2280
agcatgacaa agatcctgga gcccttcaag aagcagaacc ccgacatcgt gatctatcag 2340
tacatggacg acctgtacgt gggcagcgac ctggagatcg gccagcacag gaccaagatc 2400
gaggagctga ggcagcacct gctgaggtgg ggcctgacca cccccgacaa gaagcaccag 2460
aaggagcccc cattcctgtg gatgggctac gagctgcacc ccgacaagtg gaccgtgcag 2520
cccatcgtgc tgcccgagaa ggacagctgg accgtgaacg acattcagaa gctggtgggc 2580
aagctgaact gggccagcca gatctacccc ggcatcaagg tgaggcagct gtgcaagctg 2640
ctgaggggca caaaggctct gaccgaggtg atcccctga ccgaggaggc cgagctggag 2700
ctggccgaga cagggagat cctgaaggag cccgtgcacg gcgtgtacta cgaccccagc 2760
aaggacctga tcgccgagat ccagaagcag ggccaggcc agtggaccta ccagatctac 2820
caggagccct tcaagaacct gaagaccggc aagtacgccc gcatgcgcgg cgcccacacc 2880
aacgacgtga agcagctgac cgaggccgtg cagaagatca ccaccgagag catcgtgatc 2940
tggggcaaga ccccaagtt caagctgccc atccagaagg agacctggga gacctggtgg 3000
accgagtact ggcaggccac ctggattccc gagtgggagt cgtgaacac ccctccctg 3060
gtgaagctgt ggtatcagct ggagaaggag cccatcgtgg gcgccgagac cttctacgtg 3120
gacggcgccg ccaacaggga gaccaagctg ggcaaggccg gctacgtgac caacaagggc 3180
cgccagaagg tggtgcccct gaccaacacc accaaccaga gaccgagct gcaggctatc 3240
tacctggccc tgcaggactc aggcctggag gtgaacatcg tgaccgacag ccagtacgcc 3300
ctgggcatca tccaggccca gcccgacaag agcgagagcg agctggtgaa ccagatcatc 3360
gagcagctga tcaagaagga aggtgtac ctggcctggg tgcccgccca aagggcatc 3420
ggcggcaacg agcaggtgga caagctggtg agcgccggca tcaggaagat cctgttcctg 3480
gacggcatcg acaaggccca ggacgagcac gagaagtacc acagcaactg gagggctatg 3540
gctagcgact tcaacctgcc tcccgtggtg ctaaggaga tcgtggccag ctgcgacaag 3600
tgccagctga gggcgaggc catgcacggc caggtggact gcagccccgg catctggcag 3660
ctggactgca cccacctgga gggcaaggtg atcctggtgg ccgtgcacgt ggcctccggc 3720
tacatcgagg ccgaggtgat ccccgccgag accggccagg agaccgccta cttcctgctg 3780
```

```
aagctggccg gccgctggcc cgtgaagacc atccacaccg acaacggcag caacttcacc 3840
agcgccaccg tgaaggccgc ctgctggtgg gccggcatca agcaggagtt cggcatcccc 3900
tacaacccc agtctcaggg cgtggtggag agcatgaaca aggagctgaa gaagatcatc 3960
ggccaggtga gggaccaggc cgagcacctg aagaccgccg tgcagatggc cgtgttcatc 4020
cacaacttca agaggaaggg cggcatcggc ggctacagcc ccggcgagag gatcgtggac 4080
atcatcgcca ccgacatcca gaccaaggag ctgcagaagc agatcaccaa gatccagaac 4140
ttcagggtgt actacaggga cagcaggaac cctctgtgga agggccccgc caagctgctg 4200
tggaagggcg agggcgccgt ggtgatccag gacaacagcc acatcaaggt ggtgcccagg 4260
aggaaggcca agatcatcag ggactacggc aagcagatgg ccggcgacga ctgcgtggcc 4320
tccaggcagg acgaggactg a                                          4341
```

<210> 4

<211> 3981

<212> DNA

<213> Artificial Sequence


<220>

<223> Description of Artificial Sequence: sequence with
     optimized codons


<400> 4

```
atggccgcca gggccagcgt gctgagcggc ggcgagctgg acaggtggga gaagatcagg 60
ctgaggcccg gcggcaagaa gaagtataag ctgaagcaca tcgtgtgggc cagcaggggag 120
ctggagaggt tcgccgtgaa ccccggcctg ctggagacca gcgagggctg caggcagatc 180
ctgggccagc tgcagcccag cctgcagacc ggcagcgagg agctgaggag cctgtacaac 240
accgtggcca ccctgtactg cgtgcaccag aggatcgaga tcaaggacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tccaagaaga aggcccagca ggccgccgcc 360
gacaccggcc acagcagcca ggtgagccag aactacccca tcgtgcagaa catccagggc 420
cagatggtgc accaggccat cagccccagg accctgaacg cctgggtgaa ggtggtggag 480
gagaaggcct tcagccccga ggtgatcccc atgttcagcg ccctgagcga gggagccacc 540
ccccaggacc tgaacaccat gctgaacacc gtgggcggcc accaggccgc catgcagatg 600
ctgaaggaga ccatcaacga ggaggccgcc gagtgggaca gggtgcaccc cgtgcacgcc 660
ggccccatcg cccccggcca gatgagggag ccccgcggca gcgacatcgc cggcaccacc 720
agcaccctgc aggagcagat cggctggatg accaacaacc cccccatccc cgtgggcgaa 780
atctacaaga ggtggatcat cctgggcctg aacaagatcg tgaggatgta cagccccacc 840
agcatcctgg atatcaggca gggcccccaaa gagcccttca gggactacgt ggacaggttc 900
tacaagaccc tgcgcgccga gcaggccagc caggaggtga gaactggat gaccgagacc 960
ctgctggtgc agaacgccaa ccccgactgc aagaccatcc tgaaggccct gggaccgcc 1020
gccaccctgg aggagatgat gaccgcctgc cagggcgtgg gcggccccgg ccacaaggcc 1080
agggtgctgg ccgaggccat gagccaggtg accaacaccg ccaccatcat gatgcagagg 1140
ggcaacttca ggaaccagag gaagatggtg aagtgcttca actgcggcaa ggagggccac 1200
accgccagga actgccgcgc ccccaggaag aagggctgct ggaagtgcgg caaggagggc 1260
```

```
caccagatga aggactgcac cgagaggcag gctaatttta gggaagatct ggccttccta 1320
caagggaagg ccagggaatt ttcttcagag cagaccagag ccaacagccc caccatttct 1380
tcagagcaga ccagagccaa cagccccacc agaagagagc ttcaggtctg gggtagagac 1440
aacaactccc cctcagaagc aggagccgat agacaaggaa ctgtatcctt taacttccct 1500
cagatcactc tttggcaacg acccctcgtc acaataaaga tcggtggcca gctgaaggag 1560
gccctgctgg ccaccggcgc cgacgacacc gtgctggagg agatgagcct gcccggcagg 1620
tggaagccca agatgatcgg cggcatcggc ggcttcatca aggtgaggca gtacgaccag 1680
atcctgatcg agatctgcgg ccacaaggcc atcggcaccg tgctggtggg acctacacct 1740
gtgaacatca tcggcaggaa cctgctgacc cagatcggct gcaccctgaa cttccccatc 1800
agccccatcg agaccgtgcc cgtgaagctg aagcccggca tggacggccc taaggtgaag 1860
cagtggcccc tgaccgagga gaagatcaag gccctggtgg agatctgcac cgagatggag 1920
aaggagggca agatcagcaa gatcggcccc gagaacccct acaacacccc cgtgttcgcc 1980
atcaagaaga aggacagcac caagtggagg aagctggtgg acttcaggga gctgaacaag 2040
aggacccagg acttctggga ggtgcagctg ggcatccccc accccgccgg cctgaagaag 2100
aagaagagcg tgaccgtgct ggacgtgggc gacgcctact tcagcgtgcc cctggacgag 2160
gacttcagga gtataccccc tttaagacca atgacttaca aggcagctgt agatcttagc 2220
cactttttaa aagaaaaggg gggactggaa gggctaattc actcccaaag aagacaagat 2280
atccttgatc tgtggatcta ccacacacaa ggctacttcc ctgatccaag gatgggtggc 2340
aagtggtcaa aaagtagtgt ggttggatgg cctgctgtaa gggaagaat gagacgagct 2400
gagccagcag cagatggggt gggagcagca tctcgagacc tggaaaaaca tggagcaatc 2460
acaagtagca atacagcagc taccaatgct gcttgtgcct ggctagaagc acaagaggag 2520
gaggaggtgg gttttccagt cacacctcaa gtaccattcc tgtggatggg ctacgagctg 2580
cacccccgaca agtggaccgt gcagcccatc gtgctgcccg agaaggacag ctggaccgtg 2640
aacgacattc agaagctggt gggcaagctg aactgggcca gccagatcta ccctggcatc 2700
aaggtgaggc agctgtgcaa gctgctgagg ggcacaaagg ctctgaccga ggtgatcccc 2760
ctgaccgagg aggccgagct ggagctggcc gagaacaggg agatcctgaa ggagcccgtg 2820
cacggcgtgt actacgaccc cagcaaggac ctgatcgccg agatccagaa gcagggccag 2880
ggccagtgga cctaccagat ctaccaggag cccttcaaga acctgaagac cggcaagtac 2940
gcccgcatgc gcggcgccca caccaacgac gtgaagcagc tgaccgaggc cgtgcagaag 3000
atcaccaccg agagcatcgt gatctggggc aagactccta agttcaagct gcccatccag 3060
aaggagacct gggagacctg gtggaccgag tactggcagg ccacctggat tcccgagtgg 3120
gagttcgtga cacccctcc cctggtgaag ctgtggtatc agctggagaa ggagcccatc 3180
gtgggcgccg agaccttcta cgtggacggc ccgccaaca gggagaccaa gctgggcaag 3240
gccggctacg tgaccaacaa gggccgccag aaggtggtgc ccctgaccaa caccaccaac 3300
cagaagaccg agctgcaggc tatctacctg gccctgcagg actcaggcct ggaggtgaac 3360
atcgtgacca cagccagta cgccctgggc atcatccagg cccagcccga caagagcgag 3420
agcgagctgg tgaaccagat catcgagcag ctgatcaaga aggagaaggt gtacctggcc 3480
tgggtgcccg cccacaaggg catcggcggc aacgagcagg tggacaagct ggtgagcgcc 3540
ggcatcagga agatcctgtt cctggacggc atcgacaagg cccaggacga gcacgagaag 3600
taccacagca actggagggc tatggctagc gacttcaacc tgcctcccgt ggtggctaag 3660
gagatcgtgg ccagcgcctt caccatcccc agcatcaaca cgagacccc cggcatccgc 3720
taccagtaca cgtgctgcc ccaggctgg aagggcagcc ccgccatctt ccagagcagc 3780
atgacaaaga tcctggagcc cttcaagaag cagaaccccg acatcgtgat ctatcagtac 3840
```

```
atggacgacc tgtacgtggg cagcgacctg gagatcggcc agcacaggac caagatcgag 3900
gagctgaggc agcacctgct gaggtggggc ctgaccaccc ccgacaagaa gcaccagaag 3960
gagcccccat tcctgtggta a                                            3981
```

<210> 5

<211> 1539

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: sequence with
      optimized codons

<400> 5

```
atgggcgcca gggccagcgt gctgagcggc ggcgagctgg acaggtggga gaagatcagg 60
ctgaggcccg gcggcaagaa gaagtataag ctgaagcaca tcgtgtgggc cagcagggag 120
ctggagaggt tcgccgtgaa cccccggcctg ctggagacca gcgagggctg caggcagatc 180
ctgggccagc tgcagcccag cctgcagacc ggcagcgagg agctgaggag cctgtacaac 240
accgtggcca ccctgtactg cgtgcaccag aggatcgaga tcaaggacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tccaagaaga aggcccagca ggccgccgcc 360
gacaccggcc acagcagcca ggtgagccag aactacccca tcgtgcagaa catccagggc 420
cagatggtgc accaggccat cagccccagg accctgaacg cctgggtgaa ggtggtggag 480
gagaaggcct tcagccccga ggtgatcccc atgttcagcg ccctgagcga gggagccacc 540
ccccaggacc tgaacaccat gctgaacacc gtgggcggcc accaggccgc catgcagatg 600
ctgaaggaga ccatcaacga ggaggccgcc gagtgggaca gggtgcaccc cgtgcacgcc 660
ggccccatcg ccccggcca gatgagggag ccccgcggca gcgacatcgc cggcaccacc 720
agcaccctgc aggagcagat cggctggatg accaacaacc cccccatccc cgtgggcgaa 780
atctacaaga ggtggatcat cctgggcctg aacaagatcg tgaggatgta cagcccccacc 840
agcatcctgg atatcaggca gggccccaaa gagcccttca gggactacgt ggacaggttc 900
tacaagaccc tgcgcgccga gcaggccagc caggaggtga agaactggat gaccgagacc 960
ctgctggtgc agaacgccaa ccccgactgc aagaccatcc tgaaggccct gggacccgcc 1020
gccaccctgg aggagatgat gaccgcctgc cagggcgtgg gcggccccgg ccacaaggcc 1080
agggtgctgg ccgaggccat gagccaggtg accaacaccg ccaccatcat gatgcagagg 1140
ggcaacttca ggaaccagag gaagatggtg aagtgcttca ctgcggcaa ggagggccac 1200
accgccagga actgccgcgc ccccaggaag aagggctgct ggaagtgcgg caaggagggc 1260
caccagatga aggactgcac cgagaggcag gccaacttcc tgggcaagat ctggcccagc 1320
tacaagggca ggcccggcaa cttcctgcag agcaggcccg agcccaccgc cccccccttc 1380
ctgcagagca ggcccgagcc caccgccccc cccgaggaga gcttcaggag cggcgtggag 1440
accaccaccc ctcctcagaa gcaggagccc atcgacaagg agctgtaccc cctgaccagc 1500
ctgaggagcc tgttcggcaa cgaccccagc agccagtga                         1539
```

Claims

1. Nucleic acid sequence encoding the gag and pol polypeptides, whereby the amino acid Ala is encoded by the nucleic acid triplett gcc or gct, Arg is encoded by agg or aga, Asn is encoded by aac or aat, Asp is encoded by gac or gat, Cys is encoded by tgc or tgt, Gln is encoded by cag or cat, Glu is encoded by gag or gaa, Gly is encoded by ggc or gga, His is encoded by cac or cat, Ile is encoded by atc or att, Leu is encoded by ctg or ctc, Lys is encoded by aag or aat, Met is encoded by atg, Phe is encoded by ttc or ttt, Pro is encoded by ccc or cct, Ser is encoded by agc or tcc, Thr is encoded by acc or aca, Trp is encoded by tgg, Tyr is encoded by tac or tat, Val is encoded by gtg ot gtc and the stop codon is tga or taa.

2. Nucleic acid sequence according to claim 1, whereby the nucleic acid sequence in the region in which the reading frames encoding the gag and pol polypeptides overlapp corresponds to the wildtyp nucleic acid sequence encoding the gag and pol polypeptides.

3. Nucleic acid sequence as depicted in SEQ ID NO:1 or 2.

4. Retroviral gag or gagpol based particle generated by nucleic acid molecules comprising the nucleic acid sequence according to any one of claims 1 to 3 having at least a 25 fold reduced incorporation rate of said nucleic acid molecules encoding the gag and pol polypeptides compared to retroviral particles generated by wildtyp nucleic acid sequences encoding the gag and pol polypeptides.

5. Retroviral gag or gagpol based particle produced by nucleic acid molecules comprising the nucleic acid sequence according to any one of claims 1 to 3 having at least a 100 fold reduced recombination rate between said nucleic acid molecules encoding the gag and pol polypeptides and a wildtype genome based transgene construct derived from the same or another retrovirus.

6. Retroviral packaging cell for the generation of the retroviral particles according to claims 4 or 5 transformed with nucleic acid molecules comprising the nucleic acid sequence according to any one of claims 1 to 3.

7. Nucleic acid molecules comprising the nucleic acid sequence according to any one of claims 1 to 3 for use as an active pharmaceutical substance.

8. Use of nucleic acid molecules comprising the nucleic acid sequence according to any one of claims 1 to 3 for the preparation of a pharmaceutical composition for the treatment and prophylactic of diseases caused by retroviruses.
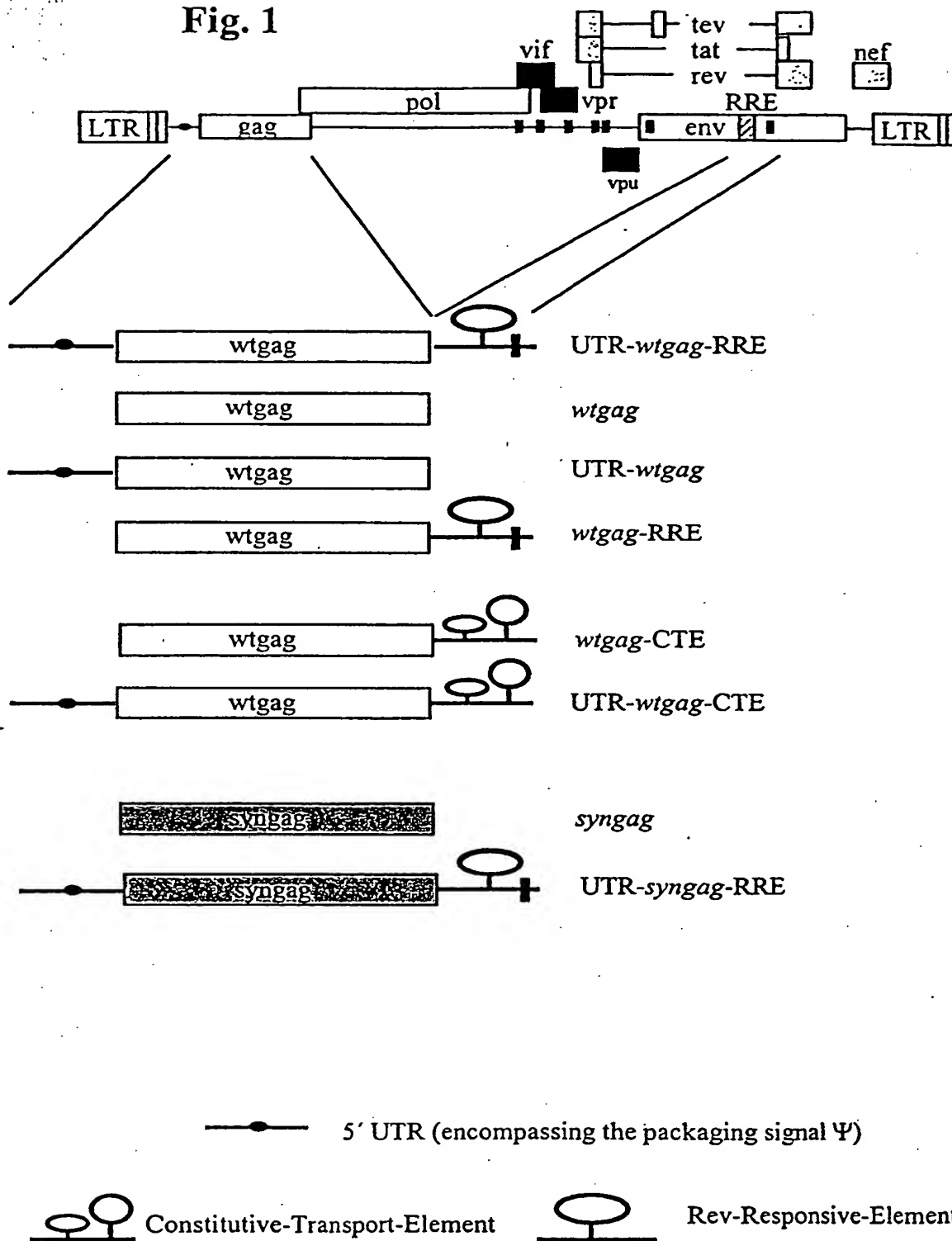
Fig. 1

5' UTR (encompassing the packaging signal Ψ)

Constitutive-Transport-Element          Rev-Responsive-Element

# Fig. 2A/1

```
  1 ATGGGCGCCAGGGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAGGTGGGA 50
    |||||  || || ||     ||  |  ||||||  || ||  |  || | ||||||
  1 ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGA 50

 51 GAAGATCAGGCTGAGGCCCGGCGGCAAGAAGAAGTATAAGCTGAAGCACA 100
    ||  ||  ||  | .|||||  ||  ||  ||||||  || ||||||  |  || ||  |
 51 AAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATA 100

101 TCGTGTGGGCCAGCAGGGAGCTGGAGAGGTTCGCCGTGAACCCCGGCCTG 150
    |  || ||||||  ||||||||||| ||  |  ||||||  || || ||  ||||||
101 TAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTG 150

151 CTGGAGACCAGCGAGGGCTGCAGGCAGATCCTGGGCCAGCTGCAGCCCAG 200
    |  ||  ||   || ||||||  || ||  ||||||  |||||  || ||
151 TTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATC 200

201 CCTGCAGACCGGCAGCGAGGAGCTGAGGAGCCTGTACAACACCGTGGCCA 250
    |||  |||||  ||  ||  || || ||  |  | || || ||  ||  || |
201 CCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAA 250

251 CCCTGTACTGCGTGCACCAGAGGATCGAGATCAAGGACACCAAGGAGGCC 300
    ||||  || ||  ||||  ||  || | ||||  |  || ||||||||||  |
251 CCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCT 300

301 CTGGACAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCCAGCA 350
    |  ||||||||||  ||||| ||||| |||||  || |  ||||||
301 TTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGCACAGCA 350

351 GGCCGCCGCCGACACCGGCCACAGCAGCCAGGTGAGCCAGAACTACCCCA 400
    ||  || ||  |||||  ||  ||||||||||  |||||  |||||  || |||||| |
351 AGCAGCAGCTGACACAGGACACAGCAGTCAGGTCAGCCAAAATTACCCTA 400

401 TCGTGCAGAACATCCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCAGG 450
    |  ||||||||||||||||||||||  || |||||  ||  |||||||||||    || ||
401 TAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGA 450

451 ACCCTGAACGCCTGGGTGAAGGTGGTGGAGGAGAAGGCCTTCAGCCCCGA 500
    ||    |  || ||  ||||||  ||  ||  || ||  |||||||||  ||||||||| ||
451 ACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGA 500

501 GGTGATCCCCATGTTCAGCGCCCTGAGCGAGGGAGCCACCCCCCAGGACC 550
    || ||  ||  ||||  |||||  || | |  |||||||||||||  ||||||  |||
501 AGTAATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATT 550

551 TGAACACCATGCTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATG 600
    |  |||||||||||  ||||||  ||||||  || ||  || ||  || ||||||||| |||
551 TAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATG 600

601 CTGAAGGAGACCATCAACGAGGAGGCCGCCGAGTGGGACAGGGTGCACCC 650
    |  ||  ||||||||||||||||||  |||||  || ||  || ||  ||||| ||  ||  || ||
601 TTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTACATCC 650

651 CGTGCACGCCGGCCCCATCGCCCCCGGCCAGATGAGGGAGCCCCGCGGCA 700
    |||||| ||  ||  || ||  ||  ||  ||  |||||||||||||||  || ||  |  ||  |
651 AGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAA 700

701 GCGACATCGCCGGCACCACCAGCACCCTGCAGGAGCAGATCGGCTGGATG 750
    |  ||||||  || ||  || ||  || ||  ||||||  |||||  || ||  || ||||||
701 GTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATG 750

751 ACCAACAACCCCCCCATCCCCGTGGGCGAAATCTACAAGAGGTGGATCAT 800
    ||  || ||  ||  || ||||||  || ||  ||||||  || ||  ||  ||||||  ||
751 ACAAATAATCCACCTATCCCAGTAGCAGAAATTTATAAAAGATGGATAAT 800
```

# Fig. 2A/2

```
 801  CCTGGGCCTGAACAAGATCGTGAGGATGTACAGCCCCACCAGCATCCTGG  850
      ||||||  |  ||  ||  ||  ||  ||  |||||| ||||| |||||||| ||||
 801  CCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGG  850

 851  ATATCAGGCAGGGCCCCAAAGAGCCCTTCAGGGACTACGTGGACAGGTTC  900
      | ||  ||  ||  ||  |||||| ||  ||  ||  |||||| ||  ||| ||||||
 851  ACATAAGACAAGGACCAAAAGAACCTTTTAGAGACTATGTAGACCGGTTC  900

 901  TACAAGACCCTGCGCGCCGAGCAGGCCAGCCAGGAGGTGAAGAACTGGAT  950
      ||  ||  ||  ||  |  |||||||| ||   |||||||| ||  ||  ||||||
 901  TATAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGAT  950

 951  GACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCC  1000
      |||  ||  |||  ||  |||| ||  ||  ||  |||||  ||  ||  |||||| ||
 951  GACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTT  1000

1001  TGAAGGCCCTGGGACCCGCCGCCACCCTGGAGGAGATGATGACCGCCTGC  1050
      |  ||  ||   |||| ||  ||  ||  ||  ||  |||||||| ||  ||
1001  TAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGT  1050

1051  CAGGGCGTGGGCGGCCCCGGCCACAAGGCCAGGGTGCTGGCCGAGGCCAT  1100
      ||||| ||  ||  ||  |||||||| |||||  ||  ||   |||| ||  ||  ||
1051  CAGGGAGTAGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAAT  1100

1101  GAGCCAGGTGACCAACACCGCCACCATCATGATGCAGAGGGGCAACTTCA  1150
      ||||||| ||  ||  ||  ||  ||  |||||| |||||||||||| |||||  ||  |
1101  GAGCCAAGTAACAAATACAGCTACCATAATGATGCAGAGAGGCAATTTTA  1150

1151  GGAACCAGAGGAAGATGGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCAC  1200
      |||||||  ||  ||||||||  ||||| |||||  ||  |||||  ||  ||  |||
1151  GGAACCAAAGAAAGATGGTTAAGTGTTTCAATTGTGGCAAAGAAGGCCAC  1200

1201  ACCGCCAGGAACTGCCGCGCCCCCAGGAAGAAGGGCTGCTGGAAGTGCGG  1250
      ||  |||||  ||  |||  |  ||||| ||||| ||||||||  ||||| ||  ||
1201  ACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGG  1250

1251  CAAGGAGGGCCACCAGATGAAGGACTGCACCGAGAGGCAGGCCAACTTCC  1300
      |||||  ||  ||||| ||||| ||  ||  ||  |||||  |||||  ||  ||
1251  AAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTT  1300

1301  TGGGCAAGATCTGGCCCAGCTACAAGGGCAGGCCCGGCAACTTCCTGCAG  1350
      |  ||  ||||||||||||   |||||||||| |||||  ||  ||  ||  ||  |||
1301  TAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAG  1350

1351  AGCAGGCCCGAGCCCACCGCCCCCCCCTTCCTGCAGAGCAGGCCCGAGCC  1400
      ||||| ||  |||||| ||  |||||| ||  ||  ||  |||||||||| ||  ||||||
1351  AGCAGACCAGAGCCAACAGCCCCACCATTTCTTCAGAGCAGACCAGAGCC  1400

1401  CACCGCCCCCCCCGAGGAGAGCTTCAGGAGCGGCGTGGAGACCACCACCC  1450
      ||  |||||| ||  ||  |||||||||||||||   ||  ||  |||||| ||  ||  |
1401  AACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAACTC  1450

1451  CCCCCCAGAAGCAGGAGCCCATCGACAAGGAGCTGTACCCCCTGACCAGC  1500
      ||||  ||||||||||||||||||||  ||  |||||||| |||||  ||  |  |  ||  |
1451  CCCCTCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCC  1500

1501  CTGAGGAGCCTGTTCGGCAACGACCCCAGCAGCCAGTGA  1539   SYNTHETISCH
      ||  ||  ||       ||  ||  |-||||||||||||        ||  |  |
1501  CTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAA  1539   WILDTYP
```

# Fig. 2B

|       | synthetic | wildtype |        |
|-------|-----------|----------|--------|
| A     | 23.78%    | 36.97%   |        |
| T     | 10.46%    | 18.91%   |        |
| G     | 32.62%    | 24.17%   | ds DNA |
| C     | 33.14%    | 19.95%   |        |
| A/U   | 22,7%     | 45,9%    |        |
| G/C   | 77,3%     | 54,1%    | mRNA   |
| CpG   | 5.        | 1        |        |

## Fig. 3A



## Fig. 3B

INSDOCID: <EP    115611201 I >

## Fig. 3C



## Fig. 3D

Fig. 4A

Fig. 4B

# Fig. 5A

**SgpΔ2**

Δ9500–10535

U3 R U5   gag   vif   rev1   rev2   U3 R U5

vpr

Δ2: 1167–1206   pol   vpx   env   nef

tat1   tat2   nef

Δ6860-8013

BGHpAd

**SIVgfpΔBP**
**SIVgfpΔDP**

ΔBP: 1167–1206
ΔDP: 1167–1206

Δ9500–9688   Δ9732–10231

U3 R U5   gag   vif   rev1   rev2   U3 R U5

vpr

pol   vpx   env   nef

tat1   tat2

GFP

**ViGΔBH**

Δ2110–6496

Δ9500–9688  Δ9732–10231

U3 R U5   gag   vif   rev1   rev2   U3 R U5

vpr

pol   vpx   env   nef

tat1   tat2

Δ6860-8013

SFFV-Pr   GFP

# Fig. 5B

**Hgp** syn

CMV   gag   BGHpAd

pol

**Sgp** syn

CMV   gag   BGHpAd

pol

**Fig. 6A**



**Fig. 6B**



**Fig. 6C**

Capture ELISA
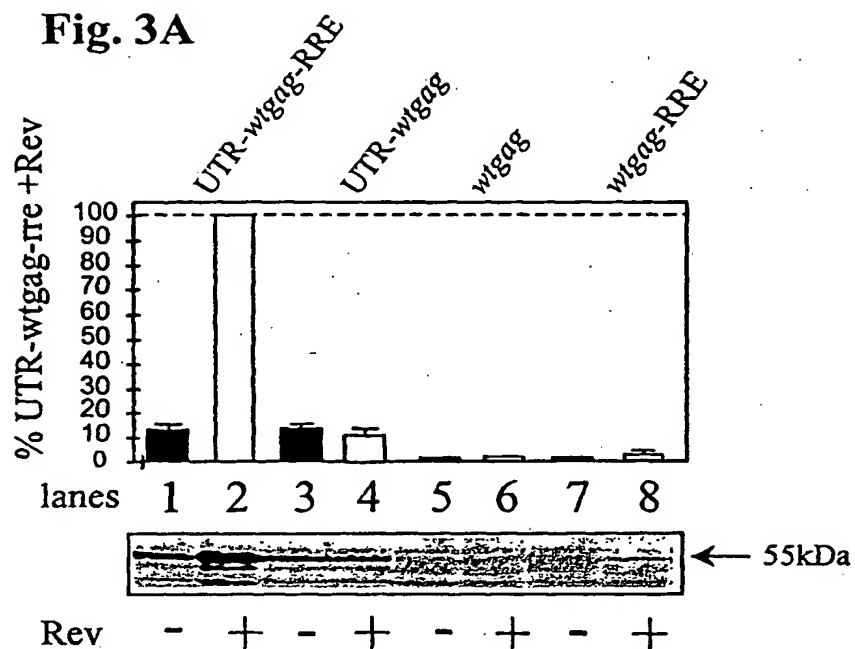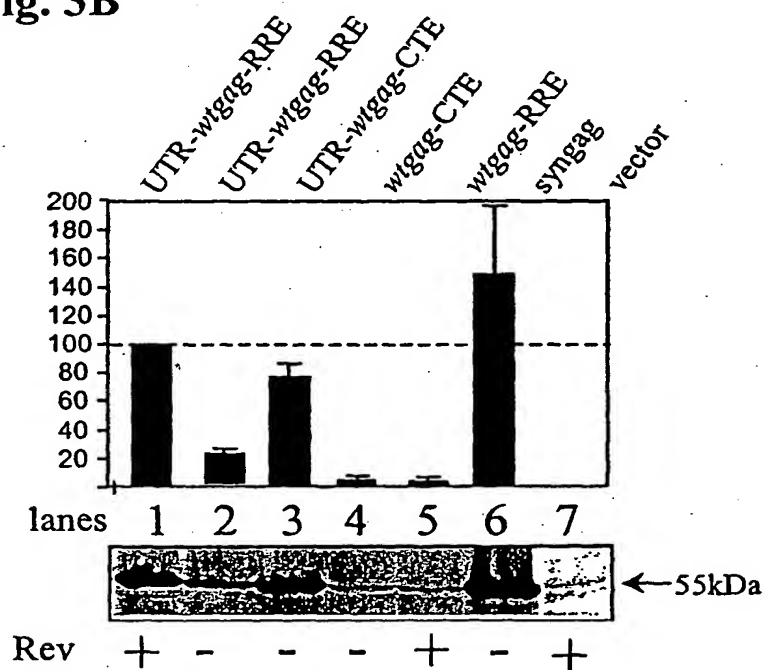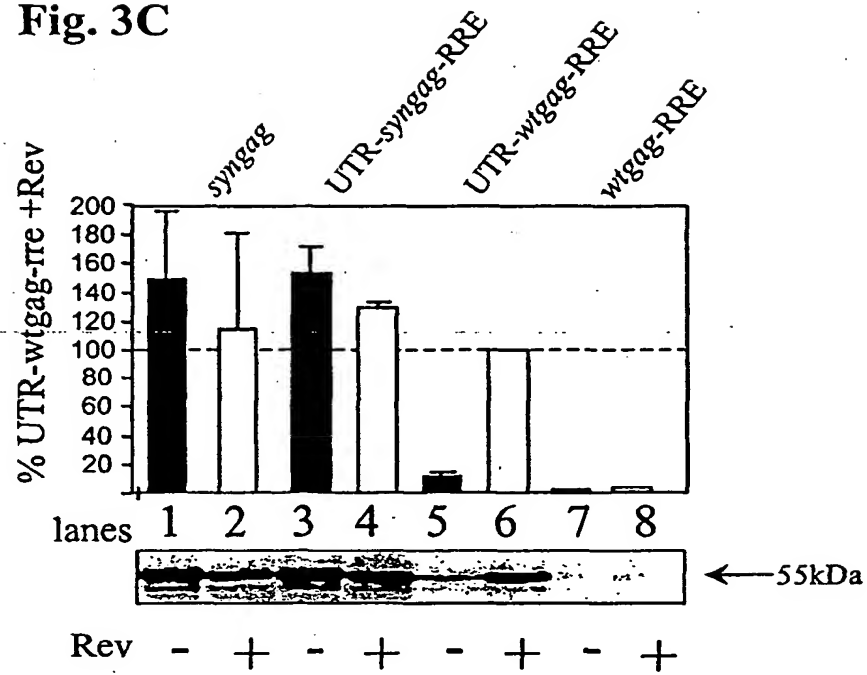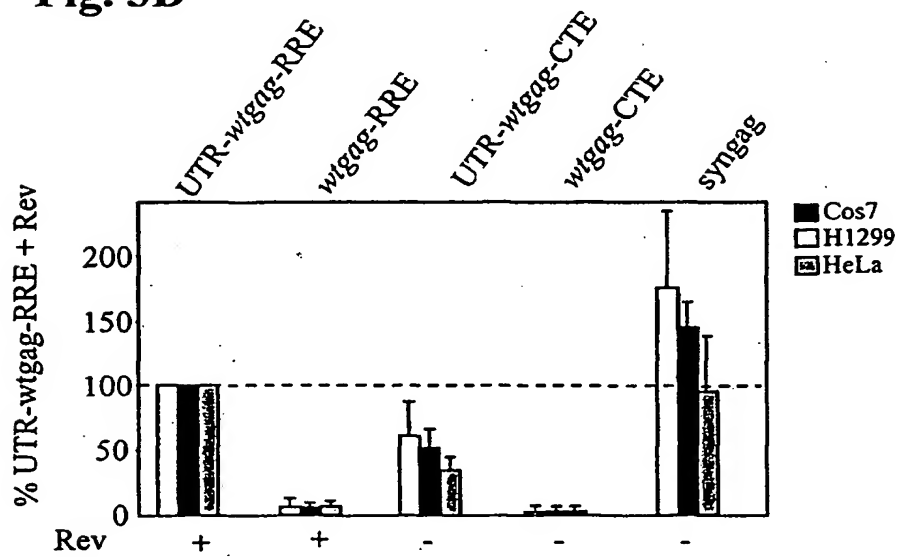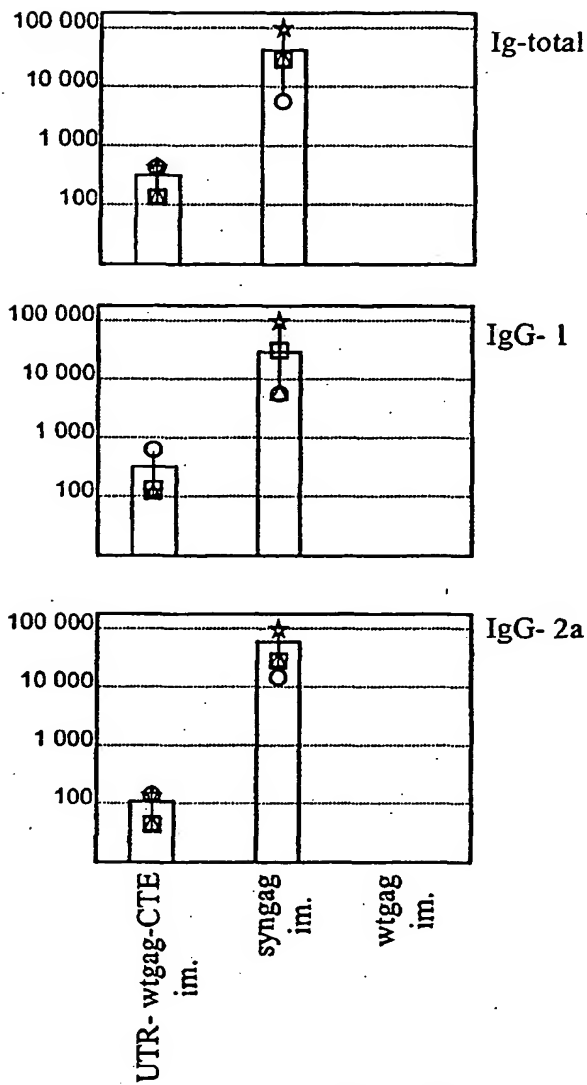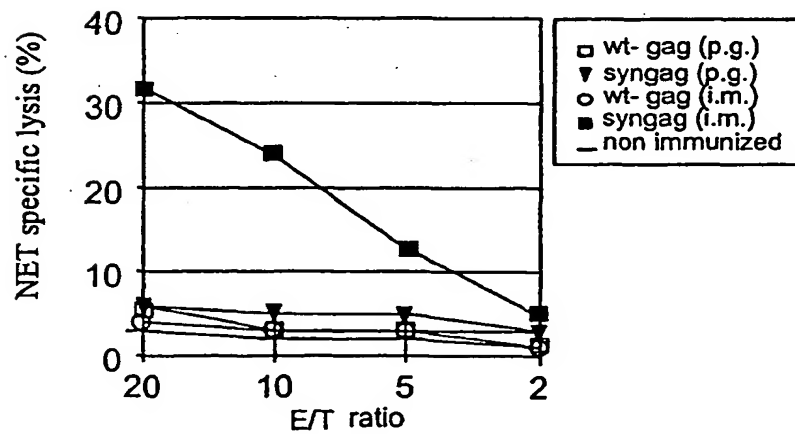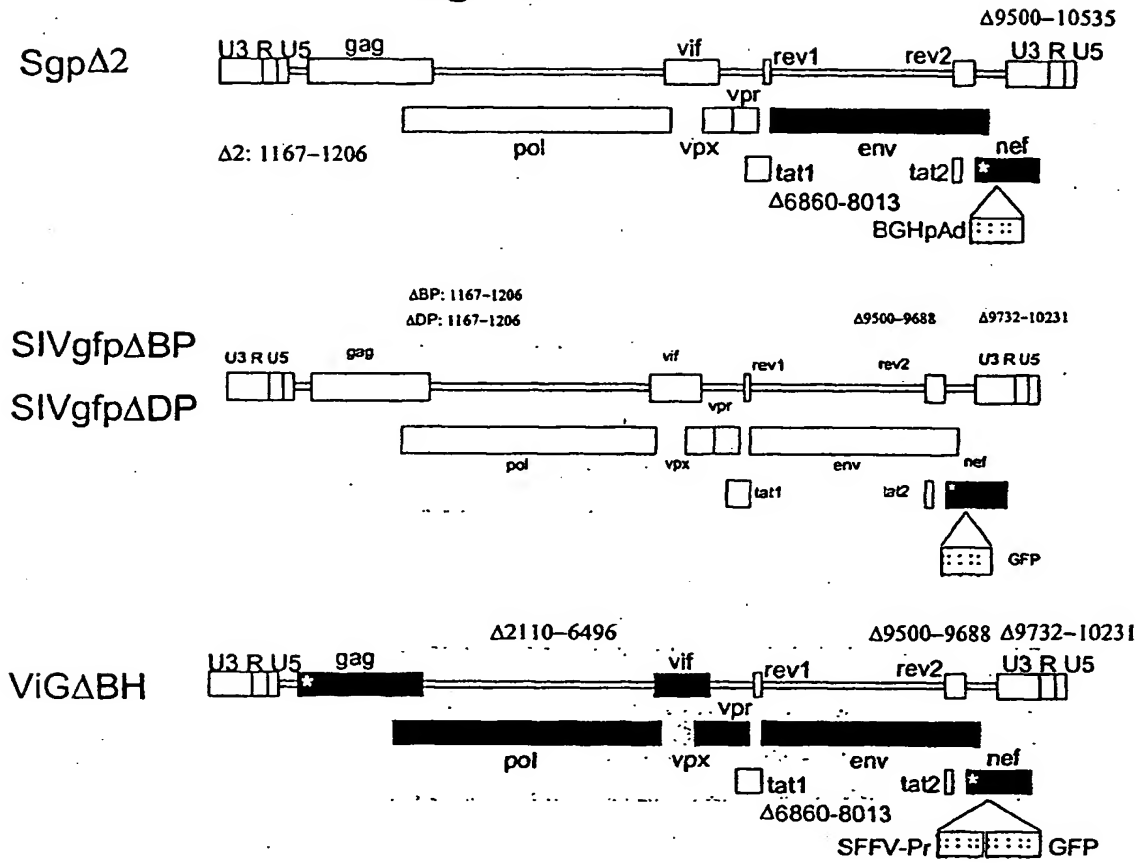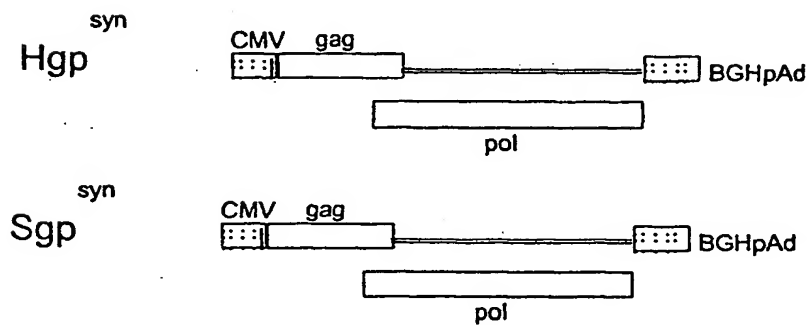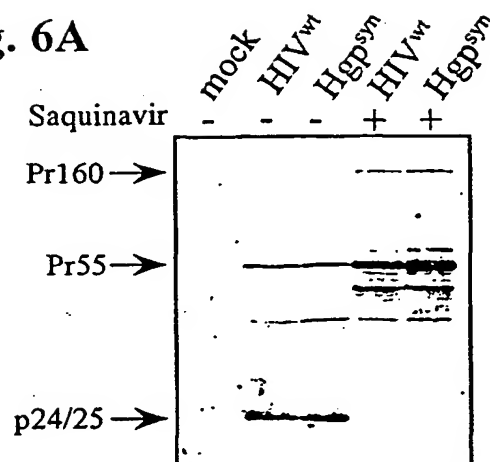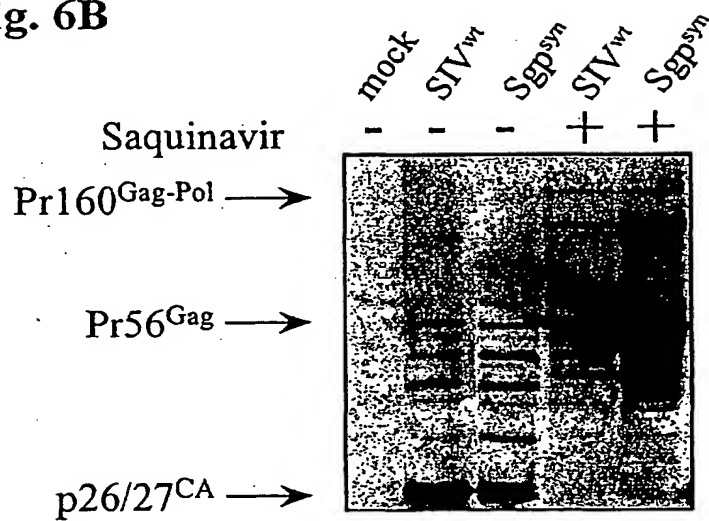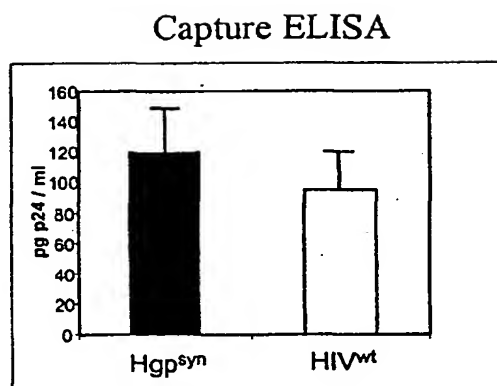
**Fig. 7**
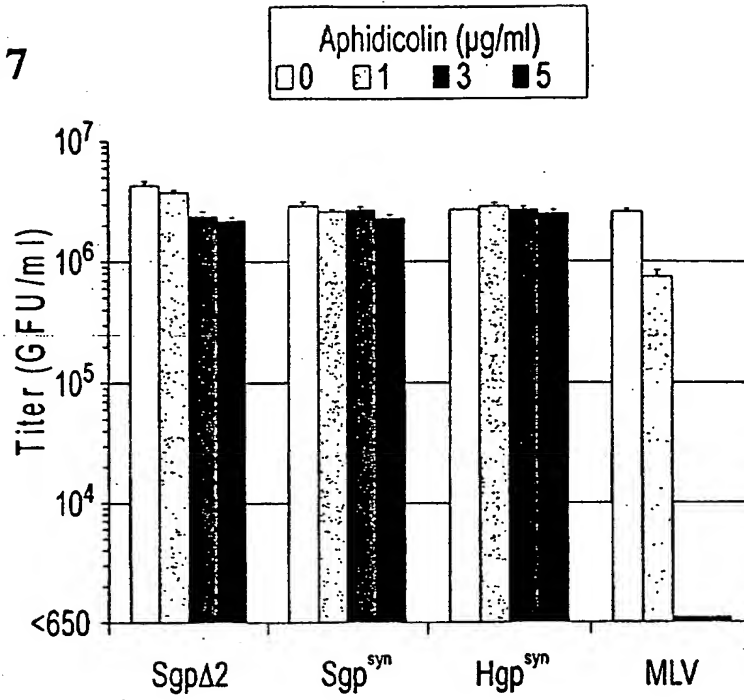


**Fig. 8A**
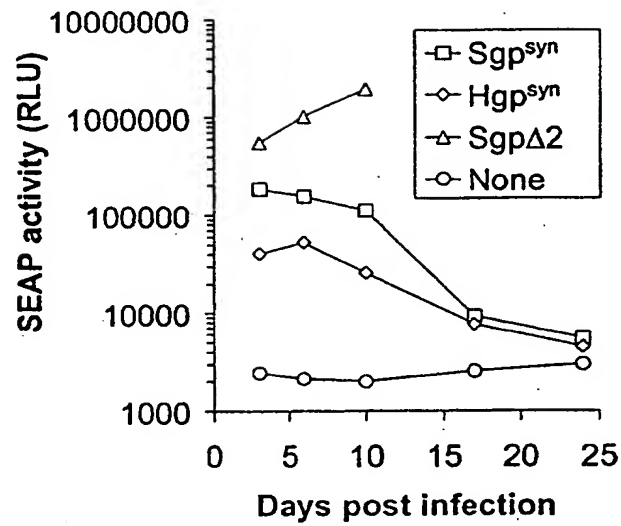
# Fig. 8B
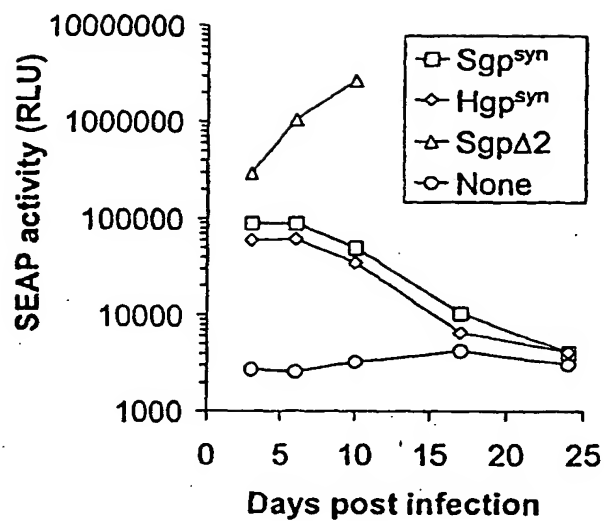


# Fig. 9
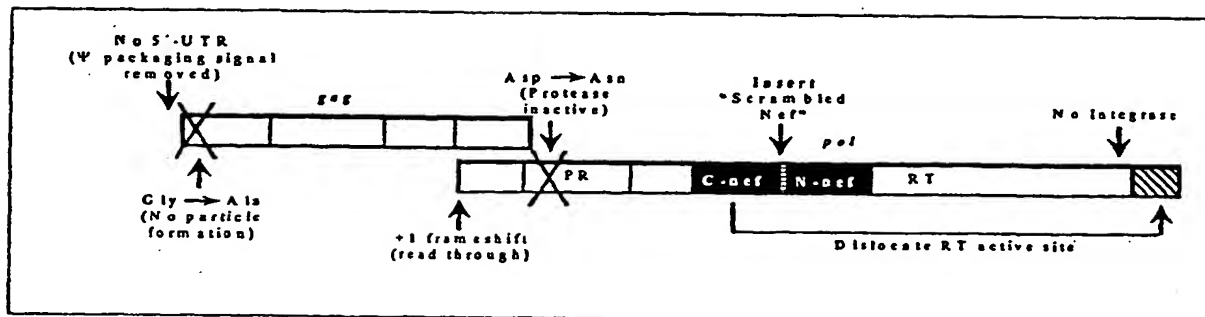
## EUROPEAN SEARCH REPORT

European Patent Office

### DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim | CLASSIFICATION OF THE APPLICATION (Int.Cl.7) |
|---|---|---|---|
| X | KOTSOPOULOU E ET AL: "A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene" JOURNAL OF VIROLOGY, vol. 74, no. 10, May 2000 (2000-05), pages 4839-4852, XP002152133 AMERICAN SOCIETY FOR MICROBIOLOGY US *whole document especially Table 1* | 1-8 | C12N15/48 C12N7/04 C12N5/10 A61K48/00 |
| X | WO 98 34640 A (SHIVER J ET AL;MERCK & CO) 13 August 1998 (1998-08-13) * examples 3-6 * | 1-8 | |
| E | WO 00 39304 A (CHIRON CORP) 6 July 2000 (2000-07-06) * page 24; claims 1-6 * | 1,2,4-8 | |

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

C07K
C12N
A61K

The present search report has been drawn up for all claims

| Place of search | Date of completion of the search | Examiner |
|---|---|---|
| THE HAGUE | 7 November 2000 | Cupido, M |

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
............................................................
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

39

## ANNEX TO THE EUROPEAN SEARCH REPORT
## ON EUROPEAN PATENT APPLICATION NO.

EP 00 11 0623

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

07-11-2000

| Patent document cited in search report | | Publication date | Patent family member(s) | | Publication date |
|---|---|---|---|---|---|
| WO 9834640 | A | 13-08-1998 | AU | 6271198 A | 26-08-1998 |
| | | | CN | 1252075 T | 03-05-2000 |
| | | | EP | 0969862 A | 12-01-2000 |
| | | | NO | 993810 A | 07-10-1999 |
| | | | PL | 335050 A | 27-03-2000 |
| WO 0039304 | A | 06-07-2000 | AU | 2221600 A | 31-07-2000 |
| | | | AU | 2487300 A | 31-07-2000 |
| | | | AU | 2596600 A | 31-07-2000 |
| | | | WO | 0039302 A | 06-07-2000 |
| | | | WO | 0039303 A | 06-07-2000 |

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82